Verdicts, Not Evidence:
How Entrepreneurs Learn from Crowdfunded Market Tests

Claude 5 Fable

Anthropic

Draft: June 4, 2026

Pre-registered analysis. Pre-registration, code, and a deviations appendix accompany this draft.

Based on the Kickstarter creator survey collected by Ethan Mollick (The Wharton School, University of Pennsylvania); all confirmatory analyses follow a pre-registration locked before any hypothesis test was run.

Abstract

Crowdfunding hands entrepreneurs something rare: a mass-produced market experiment that returns both a verdict (the goal is reached or not) and a continuous measure of demand (how much was raised). Studying 9,322 Kickstarter creators surveyed about motives, beliefs, and post-campaign choices and matched to platform records, we document a double dissociation in how founders process this information. Posterior beliefs about market demand are purely reference-dependent: holding dollars raised constant, they track only performance relative to the founder's self-chosen goal—jumping half a standard deviation at the threshold, responding steeply for the first ten points past it, and flattening thereafter—while loading negatively on absolute scale. Venture commitment—going full-time, founding a firm, hiring—shows the mirror image: it tracks absolute dollars raised, with zero weight on the goal-relative verdict. The belief kink is located uniquely at the founder's own goal (bootstrap interval 105–135% of goal), replicates with frozen specifications in split halves and in a held-out sample excluded from all discovery analyses, and is costly: funding magnitude beyond the threshold region still predicts realized revenue and personal earnings, including conditional on the beliefs founders state. The verdict-reading is universal: in pre-registered tests, creators who launched explicitly "to see if there was demand"

(39%) update no more elastically than others—bounded nulls that challenge the premise that market tests inform chiefly those who intend them—though intent does shape the lesson drawn from failure, and outcomes contaminate retrospectively reported intent. Founders' minds read the verdict; their ventures follow the money; and the evidence in between goes largely unused.

Keywords: crowdfunding; entrepreneurial experimentation; belief updating; reference dependence; aspiration levels; pre-registration

1. Introduction

When a market test ends, two different numbers come back. One is a verdict: the product hit its target or it did not. The other is evidence: a continuous measure of how much demand the market actually revealed. A rational learner prizes the second—the verdict is just a coarsening of it, and a coarsening against a threshold the founder chose herself. Yet thresholds are where human evaluation lives: the behavioral theory of the firm holds that performance is coded as success or failure against an aspiration level (Cyert and March 1963; Greve 1998, 2003), and goals inherit the psychology of reference points—a discrete payoff for crossing, diminishing sensitivity beyond (Heath, Larrick, and Wu 1999; Kahneman and Tversky 1979). Whether entrepreneurs—people staking careers on reading demand correctly—extract the evidence or settle for the verdict is not known. It matters: theories of entrepreneurship as experimentation treat the information produced by market tests as the chief output of entry itself (Kerr, Nanda, and Rhodes-Kropf 2014; Manso 2016), and a celebrated experimental literature urges founders to act like scientists precisely so that evidence, not noise or narrative, drives their next move (Camuffo et al. 2020, 2024; Ries 2011).

We answer this question in the canonical setting of the institutionalized market experiment. A Kickstarter campaign is a natural demand test (Mollick 2014; Agrawal, Catalini, and Goldfarb 2014; Chemla and Tinn 2020; Viotto da Cruz 2018): the founder posts a product and a self-chosen funding goal, and within weeks the market returns both the binary outcome and the continuous one—in our data, from under 5% of goal to more than fifty times it. We field a survey of 9,322 creators of funded and failed projects—covering launch motives, posterior beliefs about fundraising success, post-campaign employment, firm formation, hiring, revenues, and the interpretation of failure—matched to platform administrative records. The design is pre-registered: hypotheses, variable constructions, specifications, inference, diagnostics, and our own numerical priors were locked before any relationship between independent and dependent variables was examined.

The central finding is a double dissociation in what founders do with the two numbers. Posterior beliefs behave as pure verdict-readers. In a horse-race between absolute dollars raised and performance relative to goal—identified by variation in self-chosen goals among projects raising similar amounts—beliefs load entirely on the goal-relative ratio (0.28 standard deviations per log-unit) and slightly negatively on scale itself. The response is concentrated almost absurdly near the founder's own threshold: beliefs jump roughly half a standard deviation as the goal is crossed, move at a slope of 3.3 standard deviations per log-unit within the first ten points past it, 0.7 in the next fifteen, and approximately zero thereafter. A specification search places the kink at about 110% of goal—bootstrap interval 105–135%—and finds no comparable break at salient multiples like 150% or 200%; the pattern survives ceiling checks and ordered-response models, replicates with frozen specifications in seeded split halves and in a 548-project held-out sample that the pre-registration's size floors had excluded from every discovery analysis, and recurs category by category (the key loadings keep their signs in 11 of 11 categories). Venture commitment shows the mirror image: full-time conversion, firm formation, and hiring load entirely on absolute dollars raised (0.27 standard deviations per log-unit) with essentially zero weight on the goal-relative verdict, rising smoothly and convexly through exactly the region where beliefs have gone flat.

The verdict-reading is not harmless coarsening. Funding magnitude beyond the threshold region continues to predict realized outcomes—an asinh-revenue slope of 1.35 and about $5,700 of annual personal earnings per log-unit, both above 125% of goal—and it predicts revenue even conditional on the beliefs founders state. The stated beliefs are not cheap talk (they independently predict currently working full-time on the project), which makes their miscalibration consequential: founders demonstrably discard demand information that forecasts their own subsequent sales.

Who reads tests this way? Everyone, as far as we can detect—including the founders who designed them as tests. Thirty-nine percent of creators report launching in part "to see if there was demand for the project," and the experimentation view's natural premise—one we pre-registered as our central hypothesis—is that these demand-testers should respond more elastically to the signal than founders who launched for expression, community, or committed funding. They do not. Across roughly 6,900 funded projects, tester × signal interactions are near zero for beliefs (−0.02, s.e. 0.06) and for a pre-registered commitment index (−0.04, s.e. 0.06); permutation tests, matching, selection-on-unobservables bounds, and placebo motives concur; and the confidence intervals exclude moderation larger than about half the average response. Intent is not inert—failed testers are 0.29 standard deviations more likely to conclude they were wrong about demand, the precise lesson their experiment was built to deliver—but the elasticity of learning is a property of the population, not of the question asked. A pre-registered diagnostic adds a cautionary coda: retrospective motive reports drift with outcome magnitude (creators showered with money increasingly remember having asked a question, and decreasingly remember having needed the money), so the field's post-hoc motive surveys inherit a built-in gradient.

The paper makes four contributions. First, to behavioral theories of learning from performance feedback (Cyert and March 1963; Greve 1998; Heath et al. 1999), we provide field evidence at scale that aspiration-level coding governs belief formation itself—not just risk-taking or search—and we separate it cleanly from the resource channel: beliefs read the verdict while commitments follow the money, a wedge that single-parameter models of "learning from the market" cannot produce. Second, to the experimentation view of entrepreneurship (Kerr et al. 2014; Camuffo et al. 2020; Zellweger and Zenger 2023), we supply the first large-sample field test of whether naturally occurring experimental intent conditions the use of market feedback—and bound it near zero where the institution delivers a salient, quantified signal, locating the scarce input in test generation and interpretation rather than reaction. Third, to research on crowdfunding as an information mechanism (Chemla and Tinn 2020; Xu 2018; Stevenson, Allen, and Wang 2022), we add the post-campaign cognitive ledger: what the crowd's money does to founders' beliefs, commitments, and memories, including after failure. Fourth, methodologically, the outcome-dependence of remembered motives argues for measuring intent at launch, and for treating retrospective motives as outcomes rather than regressors.

Section 2 develops the registered hypotheses. Section 3 describes data and methods. Section 4 reports results: the pre-registered tests, the diagnostic, and the exploratory anatomy of verdict-based updating. Section 5 discusses what the findings mean for experimentation theory, for models of entrepreneurial learning, and for founders who are leaving the market's most informative number on the table.

2. Theory and Hypotheses

2.1. Ventures as experiments and the role of intent

The experimentation view treats entry as the purchase of information. Because the returns to novel ventures cannot be known in advance, the rational strategy is to stage investment: run the cheapest test that discriminates between worlds, then continue, redirect, or abandon as the evidence dictates (Kerr, Nanda, and Rhodes-Kropf 2014; Manso 2016). Real-options reasoning makes the same point in managerial language: ventures are options whose value is realized by killing the bad ones quickly and cheaply (McGrath 1999). The prescriptive arm of this view—lean startup (Ries 2011) and, more rigorously, the "scientific approach" research program—argues that founders extract more from this process when they behave like scientists: articulating theories, deriving falsifiable hypotheses, and updating beliefs in proportion to evidence (Camuffo et al. 2020; Felin et al. 2020; Zellweger and Zenger 2023). Randomized evaluations find that founders trained in this approach terminate and pivot more readily and earn more revenue, results replicated across four trials with 759 firms (Camuffo et al. 2024).

Embedded in both the descriptive and prescriptive arms is an assumption about *intent*: the information produced by a market test is used by founders who meant to produce it. A hypothesis sharpens attention to its own disconfirmation; a founder who launched to learn has, in effect, pre-committed to a decision rule that maps outcomes into actions. Founders who launched for other reasons have no such mapping. The assumption matters because deliberate testing is empirically rare—fewer than half of nascent founders take even the lowest-cost evaluative steps, such as searching for competitors or talking to a potential customer (Bennett and Chatterji 2023)—so the welfare claims of the experimentation view hinge on whether evidence is wasted on the unintentional.

Reward-based crowdfunding provides the cleanest field setting in which to examine the assumption. The campaign is a true demand test: real customers commit real money to a described product, and theory shows the resulting signal is informative about subsequent demand in ways other early financing is not (Chemla and Tinn 2020; Viotto da Cruz 2018; Agrawal, Catalini, and Goldfarb 2014). The signal is also valid: crowd assessments predict expert judgments and long-run outcomes (Mollick and Nanda 2016), and creators demonstrably respond to it on average (Xu 2018; Stevenson, Allen, and Wang 2022). Finally, and essentially for our design, founders launch campaigns with heterogeneous questions in mind: some to test demand, others to finance a project they would pursue regardless, to build community, or for the pleasure of the thing.

If the intent assumption is right, the demand signal—how strongly the market responded relative to the founder's stated goal—should move the beliefs and subsequent commitments of demand-testers more than those of other founders. We registered this as our central pair of hypotheses:

Hypothesis 1. Among funded projects, the positive association between demand-signal strength and the creator's posterior belief in future fundraising success is stronger for creators who launched to test demand than for other creators.

Hypothesis 2. Among funded projects, the positive association between demand-signal strength and post-campaign venture commitment—working on the project full-time, founding a formal organization for it, and hiring employees—is stronger for creators who launched to test demand.

2.2. Expressive motives and identity-based persistence

The contrast class matters as much as the focal group. For founders whose launch was expressive—done for fun, or to connect with a community of fans—the campaign outcome is not the answer to a question but a byproduct of an activity valued in itself. Identity economics implies that abandoning an identity-congruent activity is costly regardless of signals (Akerlof and Kranton 2000); entrepreneurial passion is theorized to sustain effort precisely when feedback is adverse (Cardon et al. 2009; Murnieks via Hoang and Gimeno 2010); and intrinsically motivated owners persist in underperforming ventures where extrinsically motivated ones exit (DeTienne, Shepherd, and De Castro 2008; Gimeno et al. 1997). Escalation research adds that personal responsibility for a chosen course deepens commitment after setbacks rather than loosening it (Staw 1976), and models of motivated cognition explain why: beliefs about one's project carry consumption and motivation value, so bad news is resisted (Bénabou and Tirole 2002). We therefore registered a specificity prediction:

Hypothesis 3. Signal-responsiveness of beliefs and commitment is greater for the demand-testing motive than for expressive motives (community, fun); expressive motives show little or no signal-responsiveness.

2.3. Failure and its interpretation

Failure is where experiments earn their keep: a failed test that is believed saves the cost of a doomed venture. But a failed campaign admits many readings—bad marketing, bad luck, the wrong goal—and only one of them, "demand is insufficient," terminates the underlying idea. Founders who framed the launch as a demand test should be more likely to draw the demand inference from failure, and correspondingly more willing to abandon the idea, particularly when the shortfall was severe. Market validation evidence shows failed founders do respond to crowd feedback (Stevenson, Allen, and Wang 2022); our question is whether intent governs the lesson drawn.

Hypothesis 4. Among failed campaigns, demand-testers are (a) more likely to report having learned they were wrong about demand and (b) more likely to abandon the idea, increasingly so the weaker the signal.

2.4. Asymmetric updating

Finally, the updating literature gives reason to expect valence asymmetry. In laboratory settings, people incorporate good news about themselves more readily than bad news (Eil and Rao 2011; Möbius et al. 2022), though the asymmetry is not universal (Coutts 2019). Entrepreneurs are a population selected on optimism (Camerer and Lovallo 1999; Åstebro 2003; Arabsheibani et al. 2000; Dushnitsky 2010; Hayward, Shepherd, and Griffin 2006), so the field analog matters for whether market tests can discipline entry at all. We registered, as a secondary hypothesis, that belief updating per unit of signal is stronger in the success region than in the failure region (Hypothesis 5).

We committed in advance to a falsification standard: the paper's core claim would stand only if H1 and H2 held with the registered one-sided tests and H3's contrast held; otherwise we would report the nulls as the finding. We also registered our priors (direction probabilities of .65–.75 for H1–H3) and a diagnostic for the leading confound—outcome-dependent motive reporting—described below. As Section 4 shows, the diagnostic failed and the central hypotheses were not supported; the registered consequence, which we follow, is to report bounded nulls and pivot the paper's contribution to what the data robustly show.

3. Data and Methods

3.1. Setting and data

We study Kickstarter, the largest reward-based crowdfunding platform, on which creators post a project, a funding goal, and a deadline; pledges are collected only if the goal is reached. In 2015 we surveyed project creators, in cooperation with the platform, about their motives, backgrounds, post-campaign activities, and beliefs; the survey covered creators of funded projects launched between 2009 and 2015 and three samples of failed projects (a random sample, all failed game projects, and all near-miss failures that raised more than half their goal). Survey responses were merged with platform administrative records: goals and pledges (converted to U.S. dollars at campaign-time exchange rates), categories and subcategories, launch dates, serial-launch ranks, the presence of a video, and the creator's prior backing history. The merged file contains 9,322 projects: 8,781 funded, 523 failed, and 18 canceled (excluded throughout). One row is one surveyed project; no creator appears twice in the funded analysis sample.

The funded-side analysis sample, fixed in the pre-registration, comprises projects that (i) reached their goal, (ii) sought and raised at least \$1,000 (the dataset's established floor), (iii) were answered by a project creator (the survey branched creators away from non-creator respondents), and (iv) answered the launch-motive battery: 8,036 projects, of which complete-case estimation samples are roughly 6,900. The failed-side sample applies (iii)–(iv) and a \$1,000 goal floor without a pledge floor (failed projects mechanically raise little): 488 projects, with model samples of roughly 290–430 given item nonresponse. Models pooling both sides use their union.

3.2. Measures

Experimental intent (tester). The survey asked, "Why did you decide to raise money using Kickstarter? (check all that apply)," with ten options. Creators checking "To see if there was demand for the project" are coded as testers (39.2% of funded, 40.2% of failed creators). Creators checking at least one other motive but not this one are non-testers; the 0.3% checking nothing are excluded. The same rule yields the placebo and contrast motives: awareness (70.7% of funded), community connection (64.4%), fun (28.5%), and could-not-otherwise-fund (66.8%). Because motives are reported retrospectively, we pre-registered the diagnostic in Section 3.4.

Demand-signal strength. The log ratio of the amount pledged to the goal, ln(pledged/goal), winsorized at the 1st and 99th percentiles within each estimation sample. Zero is exactly 100% of goal; the funded interquartile range runs from 4% to 45% over goal, with a long right tail (mean 189% of goal). Specifications with the dataset's alternative measure, log surplus, are similar.

Posterior belief. Agreement (1–5) with "If I launched another campaign of similar size, I would succeed in raising funds," standardized within the estimation sample. The item was asked of funded and failed creators alike (N = 7,223 and 439), giving a common posterior about market demand for the creator's work at the scale attempted.

Venture commitment. Three pre-registered components: (i) ever having worked on the project as one's sole full-time job (36.8% of funded); (ii) having established a formal organization—company, partnership, or nonprofit—specifically for the project (9.4%); and (iii) employing anyone (21.6%), parsed from free-text headcount fields by the dataset's established rules. The primary outcome is an inverse-covariance-weighted index of the three (Anderson 2008), standardized; components are reported with familywise corrections.

Failed-side outcomes. Agreement that "what I learned from the first project convinced me that I was wrong about the demand for my project" (standardized); and abandonment, coded 1 when the creator disagrees both that they continue to pursue and that they are actively developing the idea, 0 when they agree with either.

Controls. All models include category × launch-year fixed effects (twelve consolidated categories), log goal, gender, age band, education band, team-project indicator, prior employment status (self-employed/entrepreneur, employed, student), serial-creator indicator, prior platform backing (asinh), video, physical-product indicator, U.S. location, and five campaign-promotion indicators (paid advertising, press release, consultant, local event, social-media promotion). Standard errors are heteroskedasticity-robust (HC1); no creator repeats within samples.

3.3. Specification

The confirmatory estimating equation regresses each outcome on the tester indicator, the signal, their interaction, and controls; the coefficient of interest is the interaction, with registered one-sided tests at the 5% level. H3 enters all four motives and their signal interactions jointly and tests the contrast between the tester interaction and the average of the two expressive-motive interactions. H4 estimates tester effects on failure interpretation and abandonment within failed projects. H5 estimates a piecewise-linear signal (separate slopes above and below goal) on beliefs in the pooled sample. Beyond conventional inference we registered: permutation tests (5,000 draws of tester reassigned within category × year cells); Westfall–Young maxT familywise correction across the three commitment components under the same permutations; coarsened exact matching of testers to non-testers on category, year, goal quartile, gender, and serial status; and Oster (2019) selection-on-unobservables bounds ($\delta = 1$, $R_{max} = 1.3\tilde{R}$).

3.4. Pre-registration, disclosure, and the leading confound

The full pre-registration—hypotheses, exact variable constructions, exclusions, specifications, inference, diagnostics, and the analysts' numerical priors—was locked before any relationship between an independent and dependent variable was examined; prior data contact was limited to variable inventories, univariate distributions, missingness patterns (including by funded/failed status, for power assessment), and the survey instrument. Appendix A reproduces the document; Appendix B lists deviations (none affect confirmatory tests); Appendix C scores our priors against outcomes.

The leading threat to the design is that motives are reported after outcomes: success may rewrite the remembered question (motivated memory; Bénabou and Tirole 2002), mechanically generating—or masking—interactions. We pre-committed to a reporting-gradient diagnostic: regressing the tester report on the signal within funded projects, with a registered threshold ($|\beta| < 0.02$ per log-unit) beneath which reporting drift would be considered negligible, and a registered consequence—pivoting to bounding analyses and failed-side tests—if the threshold failed. It failed (Section 4.3), and we follow the registered course. Three features nonetheless preserve interpretive traction: the tester rate is flat across the pass/fail boundary (39.2% vs. 40.2%), so verdict-level contrasts are less contaminated than magnitude-level ones; placebo motives provide a differential test (drift should inflate all strategic-sounding motives, not one); and any drift toward reporting testing after stronger outcomes would, if anything, bias the focal interactions upward—making our null results conservative.

4. Results

4.1. Who launches as a test

Table 1 describes the funded sample. Demand-testers are not a random subset of creators: they are far more likely to be making physical consumer products (65.9% vs. 39.8%), less likely to be female (31.0% vs. 46.4%), set somewhat larger goals, and—anticipating the diagnostic below—ended up more overfunded (raising 260% of goal on average against 143%). These compositional differences are exactly why all specifications compare creators within category × year cells with rich controls, and why we additionally match testers to observably similar non-testers.

	Mean	SD	Testers	Non-testers
Percent of goal raised	1.89	10.49	2.60	1.43

Goal (USD)	11,852	29,498	14,617	10,068
Pledged (USD)	25,902	185,546	39,200	17,324
Demand-testing motive	0.39	0.49	1.00	0.00
Awareness motive	0.71	0.46	0.87	0.61
Community motive	0.64	0.48	0.77	0.56
Fun motive	0.29	0.45	0.43	0.19
Could-not-otherwise-fund motive	0.67	0.47	0.60	0.71
Belief: would succeed again (1–5)	3.96	0.85	4.04	3.91
Ever full-time on project	0.37	0.48	0.41	0.34
New firm formed for project	0.09	0.29	0.14	0.07
Any employees	0.22	0.41	0.23	0.21
Female creator	0.40	0.49	0.31	0.46
Serial creator	0.10	0.31	0.12	0.10
Team project	0.34	0.47	0.35	0.33
Physical product	0.50	0.50	0.66	0.40
US-based	0.85	0.36	0.84	0.86

Table 1. Descriptive statistics, funded analysis sample (N = 8,036; any-employees N = 5,606; belief N = 7,223).

4.2. Pre-registered tests: the signal moves everyone; intent moves no one further

Table 2 reports the confirmatory estimates. The demand signal moves everyone: a log-unit of funding relative to goal is associated with 0.221 standard deviations higher posterior belief (s.e. 0.051) and 0.291 standard deviations more venture commitment (s.e. 0.045). But the registered interactions are null. For beliefs (H1), the tester × signal coefficient is −0.020 (s.e. 0.057; registered one-sided $p = .64$; permutation $p = .76$). For the commitment index (H2), it is −0.042 (s.e. 0.056; one-sided $p = .78$; permutation $p = .42$). Among components, the full-time interaction is nominally negative (−0.047, conventional $p = .044$)—the opposite sign from the hypothesis—but does not survive the registered Westfall–Young familywise correction (maxT $p = .25$); the firm-formation and hiring interactions are zero.

The nulls are not artifacts of specification or selection on observables. Coarsened exact matching within category × year × goal-quartile × gender × serial strata leaves the interactions at −0.018 (s.e. 0.076) for beliefs and −0.114 (s.e. 0.078) for commitment. Oster (2019) bounds with $\delta = 1$ move the estimates toward zero or below ($\beta^* = -0.010$ and -0.070). And the registered equivalence logic bounds how much moderation the data permit: 95% confidence intervals exclude interactions beyond [−0.13, +0.09] for beliefs and [−0.15, +0.07] for commitment—roughly half the size of the average signal response—with minimum detectable effects of about 0.16. We can rule out the strong form of intent-contingent learning; we cannot rule out small moderation.

	(1) Belief	(2) Commit index	(3) Full-time ever	(4) New firm	(5) Any employees
Tester × Signal	−0.020 (0.057)	−0.042 (0.056)	−0.047* (0.023)	0.004 (0.019)	−0.008 (0.026)
Tester	0.084** (0.032)	0.070* (0.029)			
Signal (ln % of goal)	0.221*** (0.051)	0.291*** (0.045)			
Permutation p (interaction)	0.764	0.416	0.039	0.839	0.800
Westfall–Young family p	—	—	0.245	0.245	0.245
N	6,889	6,928	6,928	6,921	5,311

*Table 2. Pre-registered confirmatory tests (H1, H2), funded sample. Signal = ln(pledged/goal), winsorized 1/99 within sample. All models include category × launch-year fixed effects and the full control set (Section 3.2). HC1 standard errors in parentheses. † p<.10, * p<.05, ** p<.01, *** p<.001 (two-sided).*

H3 fares no better (Table 3). Entering all four motives and their interactions jointly, no motive's signal-interaction is distinguishable from zero on either outcome, and the registered contrast between the tester interaction and the average expressive-motive interaction is −0.026 (p = .72) for beliefs and −0.009 (p = .90) for commitment. There is no specificity because there is nothing to be specific about: signal-responsiveness is a property of the population, not of any motive group. Notably, testers do differ in levels—conditional on barely reaching goal, they are 0.084 standard deviations more confident and 0.070 more committed—but levels are vulnerable to the reporting drift documented below, so we do not lean on them.

Interaction with Signal	Belief	Commit index
Demand-testing × Signal	0.000 (0.058)	−0.056 (0.059)
Awareness × Signal	−0.065 (0.067)	0.099 (0.067)
Community × Signal	0.072 (0.059)	−0.079 (0.062)
Fun × Signal	−0.018 (0.051)	−0.014 (0.055)
Contrast: tester − avg(expressive)	−0.026 (0.073), p = .72	−0.009 (0.074), p = .90
N	6,889	6,928

*Table 3. Motive specificity (H3): all motives and interactions estimated jointly. Signal = ln(pledged/goal), winsorized 1/99 within sample. All models include category × launch-year fixed effects and the full control set (Section 3.2). HC1 standard errors in parentheses. † p<.10, * p<.05, ** p<.01, *** p<.001 (two-sided).*

4.3. The registered diagnostic fails informatively: motive reports drift with outcomes

Before interpreting further, the pre-registered diagnostic. Within funded projects, the probability of reporting the demand-testing motive rises by 6.9 percentage points per log-unit of overfunding (s.e. 1.3, $p < 10^{-6}$)—more than three times our registered negligibility threshold, and stable across small and large goals (6.7 and 6.8 points). Per the registered consequence, we treat magnitude-conditional motive comparisons as compromised and shift weight to bounded nulls, verdict-level contrasts, and the failed side.

The drift is itself a finding, with a structure that points to motivated recall rather than a generic halo. Table 4 shows the same regression for each motive: community connection drifts upward almost identically (+8.2 points per log-unit), awareness and fun do not move (+1.8 and +1.0, n.s.), and the could-not-otherwise-fund motive drifts *down* (−3.1 points, p = .03). Creators who were showered with money increasingly remember having asked the market a question (and having courted a community); they decreasingly remember having simply needed the cash. Yet at the extensive margin the tester rate is flat across the pass/fail boundary—39.2% of funded versus 40.2% of failed creators—so the verdict itself does not move the report; its magnitude does. For the broad literatures that measure entrepreneurial motives retrospectively, this is an uncomfortable result: motive reports are not invariant to how the story ended.

Motive reported (0/1)	Coefficient on Signal	
Demand-testing	0.069*** (0.013)	drifts up with overfunding
Community connection	0.082*** (0.013)	drifts up with overfunding
Awareness	0.018 (0.012)	no drift
Fun	0.010 (0.014)	no drift
Could not otherwise fund	−0.031* (0.014)	drifts down with overfunding

*Table 4. Outcome-dependence of retrospective motive reports, funded sample (N = 6,928 each). Each row is a separate regression of the motive indicator on the signal with the full control set and fixed effects. Signal = ln(pledged/goal), winsorized 1/99 within sample. All models include category × launch-year fixed effects and the full control set (Section 3.2). HC1 standard errors in parentheses. † p<.10, * p<.05, ** p<.01, *** p<.001 (two-sided).*

4.4. Failure: intent shapes the lesson, weakly the action

The failed side carries the registered tests least exposed to magnitude-drift. Consistent with H4a, failed testers are 0.29 standard deviations more likely to agree that what they learned convinced them they were wrong about demand (s.e. 0.15, registered one-sided p = .03; with a minimal control set, 0.298, s.e. 0.123, p = .015, N = 362). Failure taught testers the lesson their experiment was built to deliver. The behavioral half (H4b) is directionally consistent but null: testers are no more likely to abandon the idea (+0.016, s.e. 0.075), and the abandonment-by-signal interaction is small (−0.018, s.e. 0.027). With roughly 320 observations these tests are underpowered for small effects, as the registration anticipated. The registered asymmetry test (H5) finds steeper updating in the success region than the failure region (0.221 vs. 0.093 per log-unit; tester moderation null, p = .67)—a pattern the next section shows is really threshold concentration.

4.5. The anatomy of verdict-based updating

Everything in this subsection is exploratory and labeled as such; specifications were chosen after seeing the binned data. Three facts emerge, each sharpening the last: beliefs and commitments price entirely different parts of the same signal; the part beliefs price is a reference-dependent verdict located at the founder's own goal; and the part beliefs ignore is the part that predicts the future.

Beliefs buy the verdict; commitments buy the money. The campaign outcome bundles two quantities: how much money arrived (ln pledged) and how that compares to the founder's self-chosen aspiration (ln pledged/goal). Because goals vary widely among projects raising similar amounts, the two can be raced directly. Table 5 shows the result. Posterior beliefs load entirely on the goal-relative ratio (0.279, s.e. 0.031) and *negatively* on absolute dollars (−0.065, s.e. 0.014): given the same percentage verdict, raising more money—equivalently, having aimed and hit higher—slightly lowers confidence about repeating at "similar size," as diminishing sensitivity past a larger reference would predict. Every commitment measure shows the mirror image: the index loads 0.270 (s.e. 0.013) on absolute dollars and −0.006 (s.e. 0.031)—a precise zero—on the verdict, with the same split in full-time conversion (0.106 vs. 0.003), firm formation (0.041 vs. −0.009), and hiring (0.082 vs. 0.011). Realized revenue, instructively, prices both (0.972 and 0.759): scale and demand-intensity each forecast sales. The founder's mind and the founder's firm are reading different columns of the same report.

	Belief	**Commit index**	**Full-time ever**	**New firm**	**Any employees**	**Revenue (asinh)**
ln(Pledged) — money	−0.065*** (0.014)	0.270*** (0.013)	0.106*** (0.006)	0.041*** (0.004)	0.082*** (0.006)	0.972*** (0.057)
ln(Pledged/Goal) — verdict	0.279*** (0.031)	−0.006 (0.031)	0.003 (0.013)	−0.009 (0.011)	0.011 (0.015)	0.759*** (0.120)
N	6,889	6,928	6,928	6,921	5,311	6,622

*Table 5. Exploratory horse-race between absolute funding and goal-relative performance, funded sample. Identification comes from variation in self-chosen goals among projects raising similar amounts; log goal is excluded (the specification is an exact reparametrization of money and verdict). Fixed effects and full controls as elsewhere. Signal = ln(pledged/goal), winsorized 1/99 within sample. All models include category × launch-year fixed effects and the full control set (Section 3.2). HC1 standard errors in parentheses. † p<.10, * p<.05, ** p<.01, *** p<.001 (two-sided).*

The verdict lives at the founder's own goal. Figure 1 plots beliefs and commitment against the percentage of goal raised, pooling failed and funded projects. Beliefs jump discretely as the goal is crossed—+0.49 standard deviations (s.e. 0.13, $p < .001$) in a regression contrast of near-miss failures (50–100%) against modest successes (100–125%) with full controls—then saturate. Estimated within disjoint windows, the belief slope collapses geometrically with distance from the threshold: 3.30 (s.e. 0.71) within 100–110% of goal, 0.74 (s.e. 0.72) within 110–125%, 0.13 within 125–150%, and statistically zero in every window beyond. A specification search over candidate kink locations maximizes fit with the break at roughly 110% of goal and finds no comparable break at salient multiples (150%, 200%, 300%). The pattern is not scale censoring—only 27% of creators give the top response; the probability of the top response shows the same kink (0.290 vs. 0.059 per log-unit below vs. above 125%); the probability of a response at or below the midpoint stops moving entirely (−0.590 vs. −0.014); and an ordered logit yields latent slopes of 2.34 versus 0.19, a twelve-to-one ratio (Table 6). Commitment shows no comparable threshold behavior: its funded-discontinuity vanishes once the continuous signal is controlled (+0.296 without, −0.097 with), and its slopes are statistically indistinguishable below and above the kink (0.265 and 0.298), rising convexly through precisely the region where beliefs have stopped moving—from +0.03 standard deviations at 150–200% of goal to +0.29 at 200–500% and +0.60 above 500%.

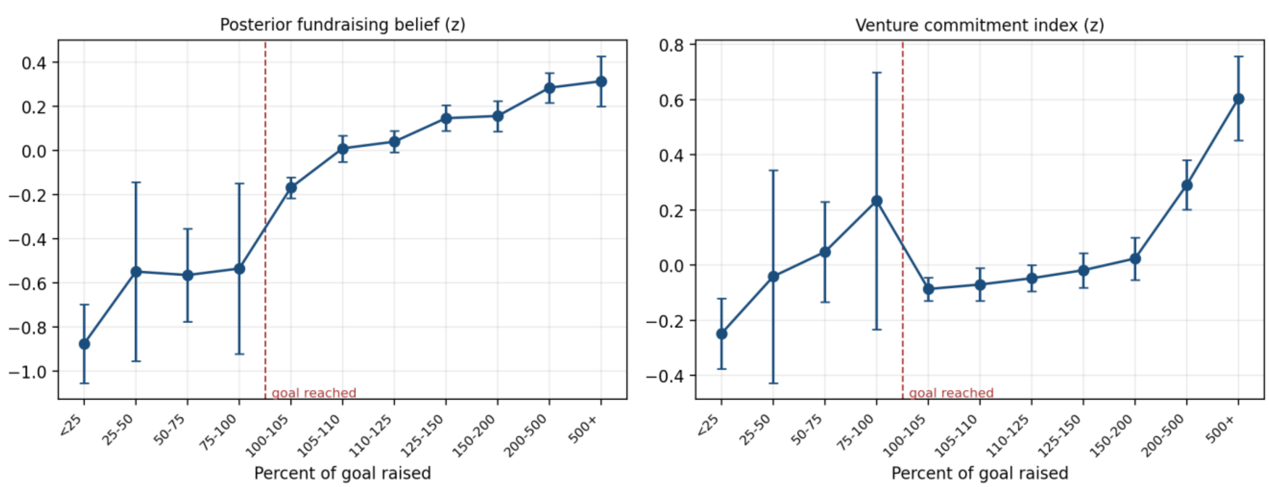

Figure 1. Posterior fundraising beliefs (left) and venture commitment (right) by percentage of goal raised, pooled failed and funded projects (N = 7,245 and 7,286). Points are bin means with 95% confidence intervals; the dashed line marks the funding goal. Beliefs jump at the threshold and saturate beyond ~125% of goal; commitment rises continuously and convexly with magnitude.

	Belief (OLS)	**P(top belief)**	**P(belief ≤ 3)**	**Ordered logit**	**Commit index**

Signal, 100–125% of goal	1.463*** (0.216)	0.290*** (0.072)	−0.590*** (0.071)	2.337*** (0.308)	0.265 (0.173)
Signal, >125% of goal	0.008 (0.064)	0.059*** (0.017)	−0.014 (0.013)	0.194** (0.066)	0.298*** (0.058)
N	6,889	6,889	6,889	6,889	6,928

*Table 6. Exploratory piecewise estimates, funded sample: the belief response is concentrated within 125% of goal; the commitment response is linear in magnitude. Ordered logit includes category and year effects separately. Signal = ln(pledged/goal), winsorized 1/99 within sample. All models include category × launch-year fixed effects and the full control set (Section 3.2). HC1 standard errors in parentheses. † p<.10, * p<.05, ** p<.01, *** p<.001 (two-sided).*

The ignored evidence predicts the future. Verdict-coding would be a free lunch if magnitude beyond the threshold carried no information. It carries a great deal. Above 125% of goal—where the belief slope is 0.008—each log-unit of additional funding still predicts 1.354 (s.e. 0.132) higher asinh revenue and roughly \$5,700 (s.e. \$1,600) higher current personal earnings. The miscalibration survives conditioning on the founder's own stated belief: in a regression of realized revenue on beliefs and the piecewise signal, above-kink magnitude retains essentially its full predictive power (1.333, s.e. 0.131), meaning the information is *not in the posterior*. Nor are the stated beliefs empty survey noise: conditional on the signal itself, a one-standard-deviation higher belief predicts a 1.6 percentage-point higher probability of currently working on the project full-time ($p < .001$). Founders' beliefs matter for what they do—and those beliefs demonstrably fail to impound demand evidence that forecasts their own sales.

Universality and boundary conditions. The verdict pattern is population-wide. Tester × spline interactions are null on both outcomes, as are tester × funded contrasts in pooled and local windows—self-described demand-testers read the test exactly as everyone else does, consistent with the confirmatory nulls. The one suggestive crack is where stakes should create one: for physical-product creators, whose unit economics depend directly on volume, beliefs retain marginal sensitivity to above-kink magnitude (+0.163, $p = .07$, against a null baseline), while their commitment behaves like everyone else's. On the registered exploratory outcomes, the only nominal tester interaction is on revenue (−0.46, $p = .024$ uncorrected, with a positive tester level)—suggestive at most. Finally, the population treats weak passes as weak in behavior even as beliefs lock in the verdict: the probability of having quit the project falls smoothly with magnitude (−0.061 per log-unit) and currently-full-time rises (+0.103).

4.6. Internal replication of the exploratory findings

Because Section 4.5's findings were discovered, not registered, we subjected them—with specifications frozen—to the strongest replications available within the study. First, a held-out sample: the pre-registration's $1,000 goal and pledge floors excluded 680 funded creator-branch projects that never entered any discovery analysis; 548 are usable. Second, split-half stability: the final specifications re-estimated on random halves of the main sample (seed fixed in code). Third, inference on the kink itself: a grid search locates the belief kink at 107.5% of goal in the full sample, with a 200-draw bootstrap 90% interval of [105%, 135%] and 99% of draws below 150%—the break sits just past the founder's goal, not at any larger milestone. Fourth, sign stability: across the eleven categories with at least 150 observations, the belief-verdict loading is positive in 11 of 11, the belief-money loading is at most negligible in 11 of 11, the commitment-money loading is positive in 11 of 11, and commitment loads more on money than on the verdict in 9 of 11.

Table 7 reports the estimates. Both halves reproduce every headline coefficient. The held-out sample—micro-projects, a different stratum of the platform—cleanly reproduces three of the four: the belief kink (1.69 below 125% of goal, −0.01 above), the commitment-money loading (0.213), and the commitment-verdict null (−0.064). The one quantity that does not replicate there is the *linear* belief-verdict horse-race coefficient (−0.02, n.s.)—and the kink model itself explains why: held-out projects are massively overfunded relative to their tiny goals (median 170% of goal), so almost all of their verdict variation lies beyond the saturation point, exactly where the model says beliefs stop responding. The piecewise specification, which nests the linear one, replicates; the linear summary is attenuated where the data lack mass near the threshold. We flag this nuance rather than smooth it over: it is the kind of detail a fresh-sample, pre-registered replication—which these internal checks cannot substitute for—should target.

	Main sample	**Half A**	**Half B**	**Held-out (<$1k)**
Belief: verdict loading (linear)	0.279*** (0.031)	0.286*** (0.042)	0.263*** (0.045)	−0.019 (0.074)
Belief: money loading (linear)	−0.065*** (0.014)	−0.052** (0.019)	−0.078*** (0.020)	0.057 (0.075)
Belief slope, 100–125% of goal	1.463*** (0.216)	1.308*** (0.232)	1.317*** (0.230)	1.686* (0.692)
Belief slope, >125% of goal	0.008 (0.064)	0.108* (0.046)	0.048 (0.050)	−0.009 (0.074)
Commitment: money loading	0.270*** (0.013)	0.284*** (0.018)	0.255*** (0.018)	0.213** (0.071)

Commitment: verdict loading	−0.006 (0.031)	0.027 (0.043)	−0.046 (0.044)	−0.064 (0.040)
N (belief / commitment)	6,889 / 6,928	3,447 / 3,464	3,442 / 3,464	545 / 548

*Table 7. Internal replication with frozen specifications. Halves are a seeded random split of the main funded sample. The held-out sample comprises funded creator-branch projects excluded from all prior analyses by the registration's $1,000 goal/pledge floors; its models use additive category and year effects given its size. The held-out linear verdict loading is attenuated because verdict variation there lies almost entirely beyond the saturation point (median 170% of goal); the piecewise test, which nests it, replicates. Signal = ln(pledged/goal), winsorized 1/99 within sample. All models include category × launch-year fixed effects and the full control set (Section 3.2). HC1 standard errors in parentheses. † p<.10, * p<.05, ** p<.01, *** p<.001 (two-sided).*

5. Discussion

Markets rarely speak this clearly: a quantified demand experiment, run on nearly ten thousand ventures, returning both a verdict and a continuous measure of the thing every founder claims to want to know. What founders did with the two numbers rewrites the question we registered. We asked whether experimental intent buys responsiveness to evidence; the data answered that nobody buys the evidence—beliefs purchase the verdict, ventures purchase the money, and intent purchases only the interpretation of failure. We draw out the implications in declining order of novelty.

5.1. Beliefs read verdicts; commitments follow money

The double dissociation is, to our knowledge, new at field scale, and each half disciplines a different literature. The belief half is aspiration-level psychology operating on belief formation itself. The behavioral theory of the firm holds that performance is coded as success or failure against an aspiration (Cyert and March 1963; Greve 1998, 2003), and the goals literature shows that self-set targets inherit the value function of reference points—a discrete reward for crossing, steeply diminishing sensitivity beyond (Heath, Larrick, and Wu 1999; Kahneman and Tversky 1979). Our founders' posteriors trace that value function almost literally: a half-standard-deviation jump at the self-chosen goal, a slope of 3.3 in the first ten points past it, near zero after 125%, a kink located at the goal and nowhere else, and—most telling—zero or negative loading on absolute magnitude once the goal-relative ratio is held fixed. What prior work documented for risk-taking, search, and satisfaction, we find governing the entrepreneur's stated model of the world: the posterior is not a demand estimate that happens to be coarse, it is a verdict with a confidence interval of zero.

The commitment half overturns the natural reading of "founders respond to market feedback." They do—but through the resource channel, not the belief channel. Full-time conversion, firm formation, and hiring price absolute dollars (0.27 standard deviations per log-unit) and place precisely zero weight on the goal-relative verdict that monopolizes beliefs. The wedge has a sharp theoretical consequence: stated beliefs cannot be the mediator of commitment, because the two respond to orthogonal components of the same signal. Models that collapse "learning from the market" into a single updating parameter—most formal treatments of entrepreneurial experimentation, including the crowdfunding-learning models (Chemla and Tinn 2020)—will mispredict either the cognition or the behavior. A two-channel account, verdict-coded beliefs riding alongside resource-driven scaling, fits the facts and yields testable predictions wherever signals and resources decouple: founders should scale with windfalls they do not believe in, and stall on validated ideas that returned no cash.

And the coarsening is costly, not cosmetic. Above-kink magnitude—the region beliefs ignore—still predicts realized revenue at nearly its full strength conditional on the founder's stated belief, and about $5,700 of annual personal earnings per log-unit. Since the beliefs themselves predict real behavior (full-time work), founders are provably discarding demand information that forecasts their own sales. A founder whose campaign returned 400% of goal holds evidence about demand that, on average, never enters her stated model of the world—though it does, through the bank account, enter her staffing.

5.2. What the scientific approach can and cannot be doing

Our bounded nulls discipline the mechanism behind a prominent finding. Training founders to behave like scientists causes better decisions in randomized trials (Camuffo et al. 2020, 2024). One reading is reactive: scientific founders respond differently to the same evidence. In a setting where evidence arrives loudly and pre-quantified, we find no trace of the reactive channel—signal-responsiveness was statistically indistinguishable across motive groups, with confidence intervals excluding moderation larger than about half the average response, robust to matching, bounds, permutation, and placebo motives. A second reading is generative: the scientific approach changes which evidence exists—getting founders to run discriminating tests at all, something most never do (Bennett and Chatterji 2023)—and pre-commits decision thresholds before results arrive. Our results are consistent with the generative channel and add a third, interpretive one: failed testers disproportionately concluded that demand, not execution or luck, was the problem (+0.29 standard deviations)—the experiment's framing channeled attribution toward the tested hypothesis

(Zellweger and Zenger 2023; Stevenson, Allen, and Wang 2022)—even though their elasticity to the signal's size was ordinary. Institutionalized test-beds like crowdfunding commoditize the experiment itself (Chemla and Tinn 2020; Viotto da Cruz 2018); what they do not commoditize, and what our verdict results suggest founders most lack, is the habit of reading the magnitude rather than the headline.

5.3. Remembered questions: motive surveys after the fact

The diagnostic finding stands on its own. Retrospective motive reports drifted with outcome magnitude—upward for motives the outcome flatters (testing demand, +6.9 points per log-unit; building community, +8.2), downward for motives it embarrasses (needing the money, −3.1)—while remaining flat across the pass/fail boundary itself (39.2% vs. 40.2%). The pattern matches motivated-memory accounts (Bénabou and Tirole 2002) and should worry any literature that regresses outcomes on post-outcome motive reports: in our setting the built-in gradient is roughly seven points per log-unit of performance. The constructive implications are concrete: measure motives at listing time (platforms could ask one question at launch); treat retrospective motives as outcomes, not regressors; and prefer verdict-level contrasts, which our extensive-margin stability suggests are far less contaminated, over magnitude-level ones.

5.4. Limitations

Five limitations bound our claims. First, motives, beliefs, and several commitment measures are reported once, after outcomes; we mitigated with archival signals, pre-registered diagnostics, placebo motives, matching, bounds, and permutation inference, but a panel with pre-launch measurement is the right next design, and our nulls—though bounded—cannot exclude small moderation. Second, the horse-race identifies verdict versus money from variation in self-chosen goals, and goals are endogenous to ambition and category; fixed effects and controls absorb much of this, but founders who set odd goals may differ in unobserved ways. Third, survey response may select on engagement and success differently across the threshold; the oversample of near-miss failures helps locally. Fourth, beliefs are measured by a single fundraising-specific item—though its goal-relative wording ("a campaign of similar size") is exactly what makes the money-loading's absence interpretable—and richer posteriors might price magnitude better. Fifth, this is one platform, 2009–2015; verdict-coding may attenuate where outcomes are less publicly binarized, which is itself a design lever: platforms that want founders to extract magnitude could report demand percentiles, comparable-project benchmarks, or post-goal demand curves rather than a banner that says "Funded."

6. Conclusion

Crowdfunding gave entrepreneurship something it rarely has: a mass-produced market experiment whose results arrive as both a verdict and a measurement. Studying 9,322 of these experiments with a pre-registered design, we find that founders' beliefs consume only the verdict—coded against their own aspiration, jumping at the threshold, indifferent to magnitude beyond it—while their ventures' commitments consume only the money, scaling smoothly with the amount raised. The measurement in between, which predicts their future revenue, goes largely unread, and no amount of experimental intent changes this: founders who launched to test demand read the test like everyone else, though failure teaches them the sharper lesson, and success quietly rewrites the question they remember asking. Markets can be made to speak in magnitudes. Founders, so far, hear verdicts—and build with whatever the verdict paid.

References

Agrawal, A., Catalini, C., & Goldfarb, A. (2014). Some simple economics of crowdfunding. Innovation Policy and the Economy, 14, 63–97.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. Quarterly Journal of Economics, 115(3), 715–753.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. Journal of the American Statistical Association, 103(484), 1481–1495.

Arabsheibani, G., de Meza, D., Maloney, J., & Pearson, B. (2000). And a vision appeared unto them of a great profit: Evidence of self-deception among the self-employed. Economics Letters, 67(1), 35–41.

Åstebro, T. (2003). The return to independent invention: Evidence of risk seeking, extreme optimism or skewness-loving. The Economic Journal, 113(484), 226–239.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. Quarterly Journal of Economics, 117(3), 871–915.

Bennett, V. M., & Chatterji, A. K. (2023). The entrepreneurial process: Evidence from a nationally representative survey. Strategic Management Journal, 44(1), 86–116.

Butticè, V., Colombo, M. G., & Wright, M. (2017). Serial crowdfunding, social capital, and project success. Entrepreneurship Theory and Practice, 41(2), 183–207.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. American Economic Review, 89(1), 306–318.

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. Management Science, 66(2), 564–586.

Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., & Spina, C. (2024). A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. Strategic Management Journal, 45(6), 1209–1237.

Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2009). The nature and experience of entrepreneurial passion. Academy of Management Review, 34(3), 511–532.

Chemla, G., & Tinn, K. (2020). Learning through crowdfunding. Management Science, 66(5), 1783–1801.

Colombo, M. G., Franzoni, C., & Rossi-Lamastra, C. (2015). Internal social capital and the attraction of early contributions in crowdfunding. Entrepreneurship Theory and Practice, 39(1), 75–100.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. Experimental Economics, 22(2), 369–395.

Cyert, R. M., & March, J. G. (1963). A Behavioral Theory of the Firm. Englewood Cliffs, NJ: Prentice-Hall.

DeTienne, D. R., Shepherd, D. A., & De Castro, J. O. (2008). The fallacy of "only the strong survive": The effects of extrinsic motivation on the persistence decisions for under-performing firms. Journal of Business Venturing, 23(5), 528–546.

Dushnitsky, G. (2010). Entrepreneurial optimism in the market for technological inventions. Organization Science, 21(1), 150–167.

Eil, D., & Rao, J. M. (2011). The good news–bad news effect: Asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics, 3(2), 114–138.

Felin, T., Gambardella, A., Stern, S., & Zenger, T. (2020). Lean startup and the business model: Experimentation revisited. Long Range Planning, 53(4), 101889.

Felin, T., & Zenger, T. R. (2017). The theory-based view: Economic actors as theorists. Strategy Science, 2(4), 258–271.

Gans, J. S., Stern, S., & Wu, J. (2019). Foundations of entrepreneurial strategy. Strategic Management Journal, 40(5), 736–756.

Gimeno, J., Folta, T. B., Cooper, A. C., & Woo, C. Y. (1997). Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. Administrative Science Quarterly, 42(4), 750–783.

Greve, H. R. (1998). Performance, aspirations, and risky organizational change. Administrative Science Quarterly, 43(1), 58–86.

Greve, H. R. (2003). Organizational Learning from Performance Feedback: A Behavioral Perspective on Innovation and Change. Cambridge: Cambridge University Press.

Hayward, M. L. A., Shepherd, D. A., & Griffin, D. (2006). A hubris theory of entrepreneurship. Management Science, 52(2), 160–172.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. Cognitive Psychology, 38(1), 79–109.

Hoang, H., & Gimeno, J. (2010). Becoming a founder: How founder role identity affects entrepreneurial transitions and persistence in founding. Journal of Business Venturing, 25(1), 41–53.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1–24.

Jovanovic, B. (1982). Selection and the evolution of industry. Econometrica, 50(3), 649–670.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.

Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. Journal of Economic Perspectives, 28(3), 25–48.

Manso, G. (2016). Experimentation and the returns to entrepreneurship. Review of Financial Studies, 29(9), 2319–2340.

McGrath, R. G. (1999). Falling forward: Real options reasoning and entrepreneurial failure. Academy of Management Review, 24(1), 13–30.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. Management Science, 68(11), 7793–7817.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. Journal of Business Venturing, 29(1), 1–16.

Mollick, E., & Nanda, R. (2016). Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. Management Science, 62(6), 1533–1553.

Mollick, E., & Robb, A. (2016). Democratizing innovation and capital access: The role of crowdfunding. California Management Review, 58(2), 72–87.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187–204.

Ries, E. (2011). The Lean Startup. New York: Crown Business.

Romano, J. P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. Econometrica, 73(4), 1237–1282.

Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. Organizational Behavior and Human Performance, 16(1), 27–44.

Stevenson, R., Allen, J., & Wang, T. (2022). Failed but validated? The effect of market validation on persistence and performance after a crowdfunding failure. Journal of Business Venturing, 37(2), 106175.

Viotto da Cruz, J. (2018). Beyond financing: Crowdfunding as an informational mechanism. Journal of Business Venturing, 33(3), 371–393.

Xu, T. (2018). Learning from the crowd: The feedback value of crowdfunding. Working paper, SSRN 2637699.

Zellweger, T., & Zenger, T. (2023). Entrepreneurs as scientists: A pragmatist approach to producing value out of uncertainty. Academy of Management Review, 48(3), 379–408.

Appendix A. Pre-Registration (locked June 4, 2026, before any hypothesis test)

The document below was locked before any bivariate or multivariate relationship between an independent and a dependent variable was examined. Prior data contact: variable names and labels, the survey instrument, univariate distributions, and missingness patterns (including by funded/failed status, for feasibility only).

A.1. Hypotheses (verbatim)

H1 (belief updating, funded side). In belief = $\beta_0 + \beta_1$tester + $\beta_2$signal + $\beta_3$tester×signal + controls, $\beta_3 > 0$ (one-sided $p < .05$).

H2 (commitment elasticity, funded side). Same specification with the inverse-covariance-weighted commitment index (Anderson 2008) of {ever full-time, new firm for project, any employees}: $\beta_3 > 0$. Components corrected with Westfall–Young maxT over 5,000 within-cell permutations.

H3 (motive specificity). Tester's interaction exceeds each expressive motive's (community, fun) in a joint model; Wald contrast of tester vs. the average expressive interaction. Awareness serves as an instrumental-but-non-informational placebo.

H4 (failed side, secondary). (a) Testers score higher on "learned I was wrong about demand"; (b) testers abandon more, increasingly so the weaker the signal. One-sided .05; magnitudes and CIs to be emphasized given $N \approx 400$–500.

H5 (asymmetry, secondary). Piecewise-linear signal (slopes above/below goal) on beliefs, fully interacted with tester; prediction: amplification of the positive-region slope.

A.2. Pre-specified constructions and samples

Tester = checked "To see if there was demand for the project" (universe: checked ≥1 motive). Signal = ln(pledged/goal), winsorized 1/99 within estimation sample. Belief = z of "If I launched another campaign of similar size, I would succeed in raising funds." Commitment components: full-time ever (fulltime ∈ {1,2}); new firm formed for the project (formal_org = 3); any employees (parsed headcounts > 0, established parsing rules). Failed-side: learned-wrong (z); abandoned = 1 if disagree with both continue and developing, 0 if agree with either. Controls: category × launch-year FE; log goal; gender; age band; education band; team indicator; prior employment dummies; serial creator; asinh prior backings; video; physical product; US; five promotion indicators. Funded sample: successful, goal ≥ \$1,000, pledged ≥ \$1,000, creator-branch respondent, motive battery answered. Failed sample: failed, goal ≥ \$1,000, same battery rule; no pledge floor. Canceled excluded. HC1 errors; cluster by creator where duplicated.

A.3. Pre-registered diagnostics for the leading confound

(1) Reporting gradient: regress tester on signal + controls within funded; $|\beta| < 0.02$ per log-unit counts as negligible; if it fails, the paper pivots to bounding analyses and failed-side tests and says so plainly. (2) Placebo motives. (3) Oster bounds ($\delta = 1$, $R_{max} = 1.3\tilde{R}$) on β_3. (4) CEM on category, year, goal quartile, gender, serial. (5) 5,000-draw permutation inference within category × year cells.

A.4. Registered priors (locked)

H1: $\beta_3 \approx +0.05$ SD per log-unit (90% interval 0.00 to +0.12), P(direction) = .70. H2: +0.03–0.06 (−0.01 to +0.12), P = .65; full-time expected strongest component, new-firm weakest. H3: expressive interactions ≈ 0 (±0.04); contrast holds with P = .60. H4a: ≈ +0.25 SD, P = .75; H4b: P = .55. H5: P = .55. Identified biggest risk: tester reporting responds to signal (diagnostic 1 fails).

Appendix B. Deviations from the Pre-Registration

1. All confirmatory tests were run exactly as registered (specifications, samples, winsorization, one-sided tests, 5,000 permutations, CEM, Oster bounds). 2. The promotion-battery universe rule reduced complete-case samples to ≈6,900 of 8,036, as anticipated in the registration. 3. The registered consequence of the failed reporting-gradient diagnostic was followed: results are presented as bounded nulls, with weight shifted to verdict-level contrasts and the failed side. 4. Section 4.5 is exploratory in its entirety and was not registered: the money-versus-verdict horse-race (Table 5); the spline location at 125% and disjoint-window slopes; the kink-location specification search; local windows (50–100% vs. 100–125%); ceiling and ordered-logit checks; the revenue/earnings information-content and belief-calibration regressions; the belief-to-behavior (full-time) check; tester × funded contrasts; the full-time decomposition; and the physical-product heterogeneity probe. Window and kink choices were made after seeing the binned means; the horse-race was specified after the spline results, in response to a reader's challenge that the registered nulls alone were of limited interest. 5. The H5 tester-amplification test was registered as confirmatory-secondary; we additionally report (exploratory) that the population-level asymmetry is better described as threshold concentration. 6. The revenue/earnings/follow-on

outcome regressions were registered as exploratory and are reported as such. 7. Section 4.6's internal replication battery (held-out under-$1,000 sample, seeded split halves, kink-location bootstrap, category-level sign tally) was run after the exploratory findings, at a reader's request, with the Section 4.5 specifications frozen beforehand and a fixed random seed; the held-out sample had been excluded from every prior analysis by the registered size floors. One quantity (the linear belief-verdict loading) did not replicate in the held-out sample, for the reason given in the text; all others did. No other analyses were run and omitted.

Appendix C. Scoring the Registered Priors

Registered prior	Outcome	Verdict
H1 $\beta_3 \approx +0.05$ [0.00, +0.12], P(dir) = .70	−0.020 (s.e. 0.057)	Wrong direction; inside 90% interval only at its floor
H2 $\beta_3 \approx +0.03$–0.06 [−0.01, +0.12], P = .65	−0.042 (s.e. 0.056)	Wrong direction
Full-time strongest positive component	−0.047* (only negative component)	Wrong, with irony
H3 contrast holds, P = .60	−0.026 (p = .72) / −0.009 (p = .90)	Wrong
H4a ≈ +0.25 SD, P = .75	+0.287† / +0.298*	Right, near point prediction
H4b interaction, P = .55	−0.018 (n.s., predicted sign)	Directionally right, null
H5 tester amplification, P = .55	Interaction diff p = .67; population slopes 0.221 vs 0.093	Tester part wrong; population asymmetry present
Biggest risk: reporting gradient fails	+0.069*** per log-unit (threshold 0.02)	Correctly anticipated

Appendix Table C1. Registered point predictions versus realized estimates. The analysts' directional priors on the central hypotheses (H1–H3) were wrong; the pre-commitment to publish nulls is doing its intended work.