

Who Trades Prediction Markets?

Neighborhood Income, Background, and Probabilistic Reasoning
Among Kalshi's Most Profitable Traders

Kalshi Research

Extended Insight

September 2026

Abstract

Prediction markets are often discussed as instruments and rarely as populations. The literature on market efficiency treats the trader base as an input whose composition is assumed rather than measured, while public commentary proceeds from a strong prior that the people profiting from these markets are finance professionals drawn from wealthy places. This paper measures the population directly, using two sources and a general-population comparison sample. The first is administrative: the residential ZIP code of every direct trader on Kalshi from January 2026 onward, crosswalked to American Community Survey median household income. The second is a survey of Kalshi's highest-PnL traders across all market categories excluding Sports and Exotics, fielded to 2,360 accounts in September 2026 with 325 surveys completed, representing a 13.8 percent response rate. We report three findings. First, the exchange is a middle-income population. The ZIP codes of Kalshi traders reflect national median income, with the modal band straddling \$80,734 and with only 6.3 percent living in ZIP codes above \$150,000; excluding sports-primary traders shifts the distribution further down, not up. Second, profitability on the exchange does not run through a financial career. Eighty-two percent of the highest-PnL traders have never worked full time at an investment bank, hedge fund, private equity firm, or proprietary trading firm, falling only marginally to approx. 76 percent of those who specialize in trading "Economics" as a category, a figure that stands in sharp contrast to any professional trading venue, where such experience is effectively universal among those earning trading profits. Seventy-nine percent attended public high school, 42 percent have worked six months or more in a customer-facing service or retail job, 65 percent describe themselves as mostly or entirely self-taught about markets and probability, and 91 percent are not full-time traders. Service-work experience is near-uniform across sixteen market categories, ranging from 33 to 53 percent, so this is a property of the whole exchange rather than of one recreational corner of it. Third, this group is distinctive in ways that are cognitive rather than professional. Benchmarked against a nationally weighted sample of 1,005 U.S. adults answering identical items, it scores a mean

of 2.42 of 4 on the Berlin Numeracy Test against the public's 0.62, achieves a perfect score roughly fifteen times as often, and resists the gambler's fallacy at 93 percent against 57 percent. Where the public is nearly as confident about an unannounced Grammy nomination as about control of Congress, confidence for the group of traders falls fourfold between the two. Within the survey cohort, research intensity, distance from 50 on lopsided questions, and the ability to discriminate knowable from unknowable questions all rise broadly with numeracy. This suggests, in line with the literature, that trading success is at least in part a product of reasoning style rather exclusively professional pedigree.

Keywords: Prediction Markets, Trader Composition, Numeracy, Probabilistic Reasoning, Kalshi

1 Introduction

The case for prediction markets is a case about aggregation. Following Hayek’s (1945) argument that markets are uniquely effective at soliciting information held in fragments across many individuals, the standard defence of an event contract is that it compels heterogeneous participants to express private beliefs as a single number, with money disciplining the expression (Wolfers and Zitzewitz, 2004). The empirical literature has tested this claim where it is easiest to observe, in prices. Our own price-side work finds that quoted probabilities are well calibrated and that calibration improves with both dollar volume and unique trader count (Kagan and Baiocchi, 2026).

What price data alone cannot reveal, however, is who sets the prices. Composition is treated as an assumption: a large and diverse pool, some of it skilled. The assumption is not innocuous, because the two standard accounts of why prediction markets work imply very different populations. One holds that accuracy emerges broadly, from the ordinary operation of a heterogeneous crowd. The other holds that it is concentrated in a small set of unusually skilled or unusually well-resourced participants, the Tetlock (2005) superforecasters or their financial-professional analogue. Cowgill and Zitzewitz (2015), studying corporate markets at Google and Ford, show that individual forecast accuracy varies systematically with trader characteristics, which establishes that who participates is not incidental to what prices do.

Public discussion has already settled this question in favour of a strong version of the second account. The presumption is that the people making money on prediction markets are finance professionals in expensive cities. That presumption bears on regulation, on product design, and on the industry’s social licence, and it has not previously been tested against data.

This paper tests it with two complementary sources, and benchmarks the second against a third. The first is administrative and covers the whole exchange: the residential ZIP code of every direct Kalshi trader active from January 2026 onward, matched to American Community Survey estimates of median household income. This yields a census-like picture of where the trading population lives, with no survey nonresponse. The second is a survey of the highest-PnL traders on the platform across every category except Sports and Exotics, which lets us ask whether the population that actually profits differs from the population that participates. Finally, the survey’s numeracy, forecasting, and confidence items were fielded in identical form to a nationally weighted general-population sample, which lets us ask how far the reasoning of the profitable cohort departs from that of the country.

We find that the exchange represents a middle-income population centred on the national median. The most profitable traders come from ordinary schools, ordinary jobs, and ordinary places, they overwhelmingly taught themselves, and four in five reached the top of the profit distribution without ever having worked in the financial industry. Where they are unusual is in measured numeracy and in the handling of uncertainty, and those differences are large, graded, and internally consistent.

The remainder proceeds as follows. Section 2 situates the study and describes the exchange. Section 3 sets out the three data sources, the definitions used throughout, and the external national benchmarks against which we compare. Section 4 reports the exchange-wide neighborhood-income results. Section 5 reports the background composition of the top-PnL cohort. Section 6 reports numeracy and reasoning. Section 7 reports trading behaviour. Sections 8, 9, and 10 discuss, qualify, and conclude.

2 Context and Literature Review

2.1 Kalshi

Kalshi began operating in 2021, following approval from the U.S. Commodity Futures Trading Commission as a Designated Contract Market in November 2020, the same regulatory category that governs traditional futures exchanges such as the Chicago Mercantile Exchange. Its core product is the *event contract*, an Arrow–Debreu security paying \$1 if a specified outcome occurs and \$0 otherwise, quoted in cents between 1 and 99, with trades matched between a Maker who posts an offer and a Taker who accepts it.

Two features of the exchange make this study possible. Because Kalshi is federally regulated, accounts are identity-verified and carry a verified residential address, which permits the ZIP-level analysis in Section 4 without relying on self-report. Because contract-level profit and loss is recorded administratively, the survey frame in Section 5 can be defined on realised PnL rather than on a proxy such as volume or self-described success.

2.2 Literature Review

This paper sits at the intersection of three literatures.

The efficiency literature traces to Hayek (1945) and is formalised for prediction markets by Wolfers and Zitzewitz (2004). Its empirical core concerns prices. Kagan and Baiocchi (2026) report that Kalshi’s prices are well calibrated near resolution and that calibration improves with participation, while explicitly leaving open whether that improvement is supplied broadly or by a concentrated set of skilled traders, since price data alone cannot identify market participants. The present paper addresses that question from the population side. Manski (2006) supplies the standing caution that a market price is not in general a sufficient statistic for participants’ average belief absent assumptions about who trades and how.

The numeracy literature supplies our principal non-demographic instrument. Cokely et al. (2012) developed the Berlin Numeracy Test as a short, psychometrically validated measure of statistical numeracy that discriminates well among educated samples, where older instruments saturate. Peters et al. (2006) establish that numeracy predicts the quality of decisions under risk independently of general cognitive ability, and Lipkus et al. (2001) document the broader finding that numerate respondents extract more information from probabilistic formats. Tetlock (2005) and Mellers et al. (2015) establish that forecasting skill is measurable and unevenly distributed, and that its correlates are dispositional rather than credential-based, which is the hypothesis our composition and reasoning results jointly test.

The survey-measurement literature informs how we treat elicited probabilities. Fischhoff and Bruine de Bruin (1999) show that a response of 50 on a 0–100 chance scale frequently encodes unfamiliarity rather than a considered even-odds judgment, and Manski and Molinari (2010) document heaping at multiples of ten. How both are handled is set out in the Technical Appendix. Groves (2006) frames the nonresponse problem that bounds inference from any voluntary survey, a point Section 9 returns to. Finally, Barber and Odean (2001) link overconfidence to trading intensity, which bears directly on the metacognition results in Section 6.5.

3 Data, Definitions, and Benchmarks

3.1 Exchange-Wide Population and the ZIP Crosswalk

The first dataset covers all direct traders on Kalshi active from January 2026 onward. Each account’s verified residential ZIP code is matched to the median household income of the corresponding ZIP Code Tabulation Area (ZCTA) as estimated by the American Community Survey, and accounts are then binned into the seven income bands reported in Figure 1. Accounts whose ZIP code has no matching ZCTA, which occurs for post-office-box ZIPs, point ZIPs, and ZIPs with too few residential addresses to be delineated, are retained in an explicit “unknown / unmatched” band rather than dropped.

Two properties of this measure require emphasis. First, ZCTAs are Census approximations of postal ZIP codes, not the ZIP codes themselves; the Census delineates them decennially from tabulation blocks and explicitly advises against treating them as postal geography. Second, and more importantly, this is a *neighborhood* measure, not an individual one. A trader living in a ZIP whose median household income is \$70,000 is not thereby a \$70,000 earner. The measure characterises the communities from which Kalshi draws, which is the question Section 4 asks, and it should not be read as a distribution of trader incomes. Section 9 returns to this.

3.2 The Top-PnL Survey

The second dataset is a survey of Kalshi’s highest-PnL traders. The frame was constructed administratively from realised profit and loss across all market categories *excluding Sports and Exotics*, on the grounds that sports outcomes resolve on information that arrives during the event itself and that Exotics are combination contracts whose PnL is not comparable to that of standalone markets. Professional market-makers are excluded. Invitations went to 2,360 accounts, and fieldwork ran from 17 to 23 September 2026, yielding 325 complete surveys, a response rate of 13.8 percent. Of the 427 respondents who began the survey, 100 left before finishing and 2 were screened out by quality checks. For a proportion near 50 percent, the approximate 95 percent margin of error for the full sample is ± 5.4 points.

Where subgroup results are reported, the relevant unweighted bases and margins of error are given with them. Subgroups defined by gender other than male are not reported anywhere in this paper: the sample contains 13 female and 8 other-identifying respondents, and percentages computed on such bases would not be meaningful.

3.3 Definitions

Numeracy. Statistical numeracy is measured with the four-item Berlin Numeracy Test (Cokely et al., 2012), scored 0 to 4. The instrument is used throughout as an ordered five-level classification and, where cell sizes permit, as a banner variable against which other responses are tabulated.

Service-work experience. Defined as having worked more than six months in a customer-facing service or retail job, with retail, customer support, food service, and operations given as examples. In Figure 3 this is tabulated by market category, where category membership is not exclusive: a respondent who trades Politics and Economics appears in both bases, so the category bases sum to well above 325.

Confidence discrimination. Respondents rated confidence in each of six probability estimates on a four-point scale. We define a within-respondent discrimination measure as the share expressing

confidence on the two congressional-control items, whose resolution is genuinely forecastable from public information, minus the share expressing confidence on the Grammy nomination item, which is not. Formally, for any group s (a numeracy stratum, the full cohort, or the general-public sample of Section 3.5),

$$D_s = \frac{1}{2}(C_s^{\text{House}} + C_s^{\text{Senate}}) - C_s^{\text{Grammy}},$$

where C denotes the share answering “somewhat” or “very” confident. Because the measure differences confidence across items within the same respondents, it is unaffected by individual differences in overall scale use.

3.4 National Benchmarks

Because this paper’s central claim is comparative, the external benchmarks used are stated explicitly in Table 1, with sources, rather than asserted in passing.

Quantity	National value	Source	Year
Median household income	\$80,734	ACS 5-year estimates	2020–2024
Median household income	\$83,730	CPS ASEC, <i>Income in the U.S.</i> (P60-286)	2024
Share of K–12 students in public school	≈ 90 – 91%	NCES, Condition of Education	Fall 2021
Adults 25+ with a bachelor’s degree or higher	38.6%	Census CPS educational attainment	2024
Adults 25–39 with a bachelor’s degree or higher	42.8%	Census CPS educational attainment	2024
Employed workers with a bachelor’s degree or higher	44.5%	Census CPS educational attainment	2024
Adults whose first regular job was in restaurants/foodservice	51%	National Restaurant Association	2026
Adults who have worked in restaurants at some point	$\approx 63\%$	National Restaurant Association	2026

Table 1: External national benchmarks used for comparison throughout this paper. The ACS figure is the appropriate comparator for Section 4, since the ZIP crosswalk is itself built on ACS estimates; the CPS figure is reported for completeness because it is the more widely cited headline number.

What is notable about the cohort’s service-work figure is not its size relative to the general population but its size relative to what the prevailing stereotype of a profitable prediction-market trader would predict, and its near-invariance across sixteen market categories, which is the subject of Section 5.3.

3.5 General-Population Comparison Sample

To establish whether the reasoning results in Section 6 distinguish the cohort from the general population, rather than only describing it, the survey’s reasoning items were fielded to a national general-population sample. The TrueDot omnibus completed 1,005 interviews between 21 and 23 September 2026, with estimates weighted to national demographics and 2024 presidential vote. Question wording and correct-answer rules are identical across the two surveys for the four Berlin Numeracy Test items, the self-estimate of numeracy performance, the gambler’s fallacy item, the six probability estimates, and the six confidence ratings. Trader fieldwork ran from 17 to 23 September, so the two field periods overlap. For a proportion near 50 percent, the approximate 95 percent margin of error in the public sample is ± 3.1 points before any design effect from weighting.

4 Where Kalshi Traders Live

4.1 The Exchange-Wide Income Distribution

Figure 1 reports the distribution of Kalshi’s direct trading population across ZIP-level median household income bands, for all direct traders and separately for traders whose primary activity is outside Sports.

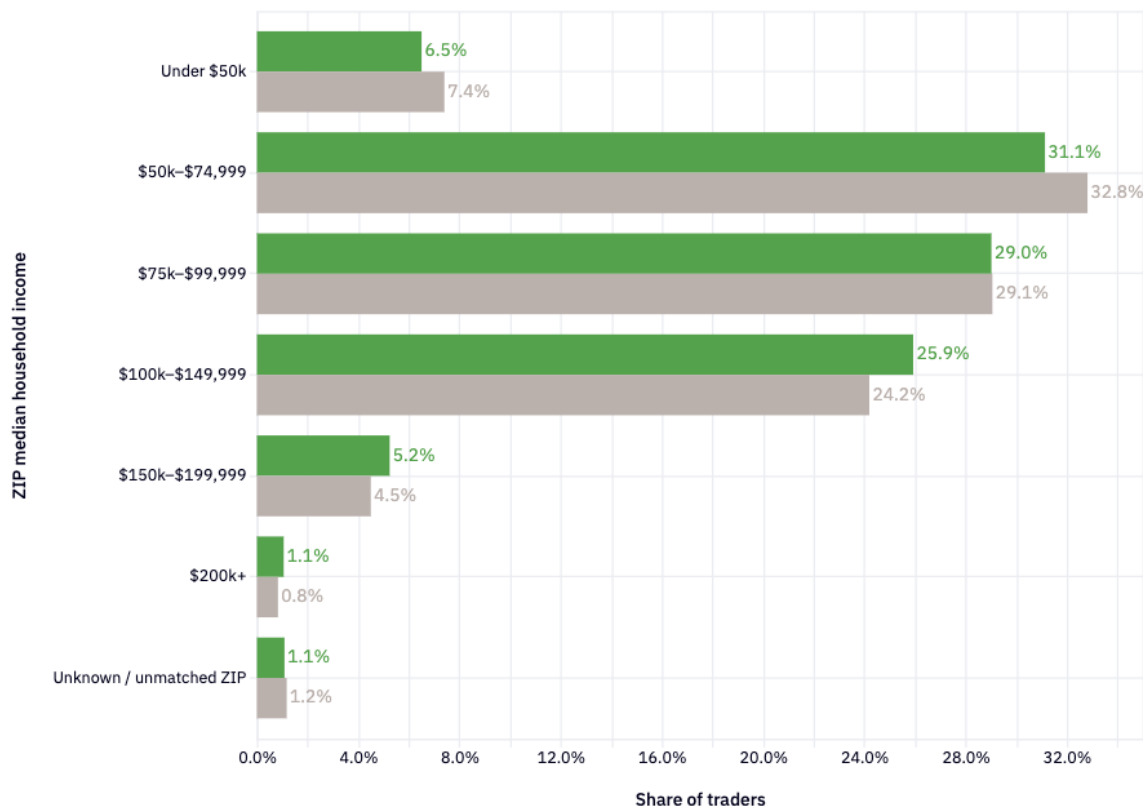


Figure 1: Distribution of Kalshi direct traders by the ACS median household income of their residential ZIP code, traders active from January 2026 onward. Green: all direct traders. Grey: traders whose primary activity is outside Sports. Bands are ZIP-level medians, not individual trader incomes.

The distribution is unimodal, centred close to the national median, and thin at the top. Two thirds of all direct traders, 66.6 percent, live in ZIP codes whose median household income is below \$100,000. The two modal bands, \$50,000–\$74,999 and \$75,000–\$99,999, together account for 60.1 percent of traders and straddle the national median household income of \$80,734. At the upper end, 5.2 percent live in ZIP codes with medians between \$150,000 and \$199,999 and 1.1 percent in ZIP codes above \$200,000, so that fewer than one trader in fifteen lives in a ZIP code above \$150,000. 1.1 percent of accounts could not be matched to a ZCTA.

4.2 Excluding Sports Moves the Distribution Down

A natural objection is that sports trading might account for the middle-income character of the exchange, with a wealthier population concentrated in the financial and political markets. Restricting to traders whose primary activity is outside Sports shifts the distribution *toward* lower-income ZIP

codes at every point of the comparison: the share below \$50,000 rises from 6.5 to 7.4 percent, the share in the \$50,000–\$74,999 band rises from 31.1 to 32.8 percent, and the combined share below \$100,000 rises from 66.6 to 69.3 percent. At the top, the share above \$150,000 falls from 6.3 to 5.3 percent.

The direction matters substantively for interpretation. If prediction markets were functioning as a financial-professional venue with a recreational sports fringe, removing the sports fringe should concentrate the remaining population income upward. It does the opposite. The non-sports core of Kalshi is, if anything, slightly less affluent in residential terms than the exchange as a whole.

5 Composition of the Top-PnL Cohort

Section 4 describes participation. This section describes profitability, using the survey of the highest-PnL traders across all categories except Sports and Exotics. Professional market-makers are excluded from the analysis.

5.1 Profiting Without a Financial Career

The single most consequential fact about this cohort is what it lacks. Eighty-two percent of the most profitable traders on Kalshi have never worked full time at an investment bank, hedge fund, private equity firm, or proprietary trading firm. Four in five of the people who consistently make money on a federally regulated derivatives exchange arrived there with no professional trading background of any kind.

The reference class that makes this figure meaningful is not the general public but the other venues in which people profit from markets. At a hedge fund or a proprietary trading desk, essentially every individual earning trading profits has a financial career by construction because the job is the qualification. Kalshi is the opposite. Its profitable cohort is overwhelmingly composed of people who acquired whatever edge they have outside the industry, and 65 percent of them describe what they know about markets and probability as mostly or entirely self-taught, against 10 percent who credit formal education.

We are precise about the comparison in the other direction as well. The 18 percent of this cohort with a financial-industry background is higher than the share of the general population with such a background, since finance and insurance together account for only around 4 to 5 percent of U.S. employment and the specific firm types named in the question are a fraction of that. Prior industry experience is therefore plainly over-represented among profitable traders relative to the country. The point is not that finance experience is rare here in absolute terms. It is that it is evidently not necessary, because the overwhelming majority of the people succeeding do not have it, and Section 5.3 shows that this holds in every corner of the exchange rather than only in the categories where one would expect amateurs to compete.

5.2 Schooling, Service Work, and Self-Teaching

Measure	Top-PnL traders	National reference
Never worked at a bank, hedge fund, PE, or prop firm	82%	near 0% at professional trading firms
Attended public high school	79%	$\approx$ 90–91% of K–12 enrollment
Attended private high school (any)	19%	$\approx$ 9–10% of K–12 enrollment
Worked 6+ months in a customer-facing service or retail job	42%	$\approx$ 63% (restaurants only); see Section 3.4
Mostly or entirely self-taught about markets and probability	65%	n/a
Mostly or entirely formally taught	10%	n/a
Family income at 16 average or below	57%	n/a
Grew up outside a large city	81%	n/a
Not a full-time trader	91%	n/a

Table 2: Background composition of Kalshi’s highest-PnL traders (base 325) against external national references. National references are constructs measured on different populations and are indicative rather than exact counterparts; see Section 3.4.

Four in five of the most profitable traders on the exchange went to public high school. Two in five have spent six months or more in customer-facing service or retail work, a rate below the national one for the same construct but far above what the prevailing image of this population would suggest. A majority grew up in households at or below average income, four in five grew up somewhere other than a large city, and only 9 percent trade full time, with 59 percent employed full time elsewhere and a further 16 percent self-employed. The modal profitable Kalshi surveyed is a trader with full-time employment (i.e. a non full-time trader), a public-school education, and no financial career, who taught themselves the subject.

Figure 2 disaggregates schooling and financial-industry background by trader category for the two categories where the comparison is most instructive, Economics and Entertainment. These sit at opposite ends of any intuitive ranking of how “financial” a market is.

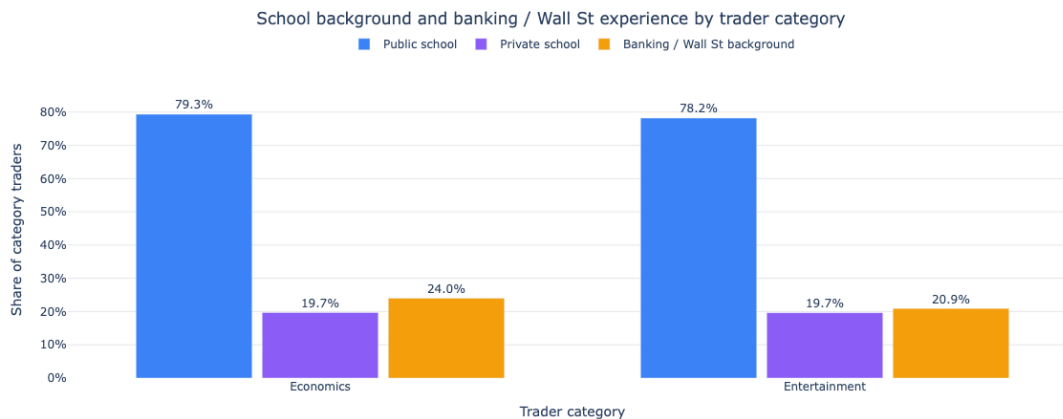


Figure 2: Public and private high school attendance and banking or Wall Street background, by trader category. Category membership is not exclusive.

The two categories are nearly indistinguishable. Public high school attendance is 79.3 percent among Economics traders and 78.2 percent among Entertainment traders; private school attendance is 19.7 percent in both. Financial-industry background differs by 3.1 points, 24.0 against 20.9

percent, which is within the error attached to subgroups of this size. Trading the Federal Reserve rather than the Grammys does not select for a different kind of schooling.

5.3 Service Work Is Near-Uniform Across Categories

Figure 3 extends the service-work measure to all sixteen market categories in which survey respondents report trading.

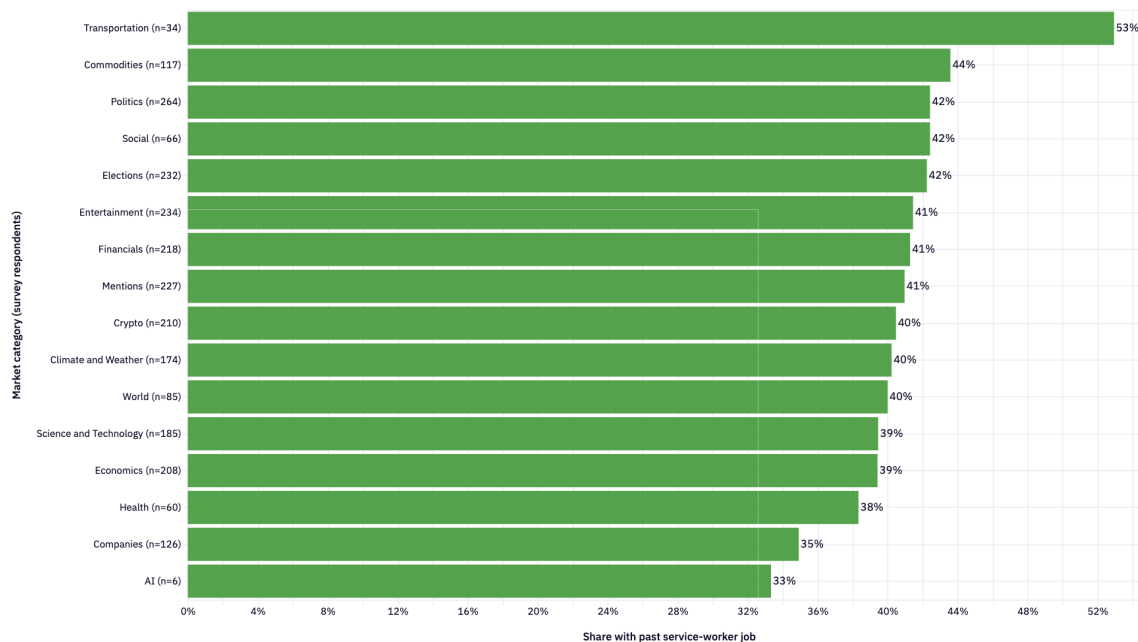


Figure 3: Share of profitable traders with 6+ months of past customer-facing service or retail work, by market category. Bases are respondents active in each category and are not mutually exclusive.

Excluding the two smallest bases, Transportation ($n = 34$) at 53 percent and AI ($n = 6$) at 33 percent, every category falls between 35 and 44 percent. Economics traders (39 percent, $n = 208$) are indistinguishable from Entertainment traders (41 percent, $n = 234$), from Elections traders (42 percent, $n = 232$), and from Crypto traders (40 percent, $n = 210$). The categories most associated with professional finance, Financials at 41 percent and Companies at 35 percent, sit inside the same narrow band as the rest.

Service-sector work history is therefore a background characteristic of roughly two in five profitable traders regardless of what they trade, which is the most direct evidence in this paper that profitability on Kalshi does not require, and does not select on, a professional financial career.

5.4 Origins and Profile

Geographic and family origins reinforce the picture. Nineteen percent of profitable traders grew up in a large urban core, 45 percent in a suburb, 14 percent in a small or mid-sized city, 14 percent in a town, and 8 percent in a rural area.

The composite is a population ordinary in origin and quantitative in training. Its members largely attended public schools, did frontline service work, grew up outside large cities in households of average means, hold other jobs, and taught themselves the subject. What they bring to the exchange is not a financial career, and Section 6 argues that what they do bring is a specific facility with numbers and uncertainty.

6 Numeracy and Probabilistic Reasoning

6.1 Level

On the four-item Berlin Numeracy Test, the cohort scores a mean of 2.42. Thirty-one percent answer all four items correctly and 54 percent answer at least three. The instrument was designed to discriminate among educated samples, where simpler numeracy scales saturate (Cokely et al., 2012). To establish what the general population scores on the same instrument, the identical four items were fielded to the nationally weighted sample described in Section 3.5; Table 4 sets the two side by side.

Berlin Numeracy Test score	0	1	2	3	4
Share of cohort	13%	17%	16%	22%	31%
Unweighted base	43	55	53	72	102
Maximum 95% margin within stratum	± 14.9	± 13.2	± 13.5	± 11.5	± 9.7

Table 3: Distribution of Berlin Numeracy Test scores, top-PnL cohort (base 325, unweighted). Mean 2.42. Margins are in percentage points for a proportion of 50 percent within each stratum, and are smaller for proportions further from 50.

	General public	Top-PnL traders
<i>Berlin Numeracy Test, number correct</i>		
Mean score (0–4)	0.62	2.42
0 correct	63%	13%
1 correct	23%	17%
2 correct	6%	16%
3 correct	6%	22%
4 correct	2%	31%
Three or four correct	8%	54%
<i>Share answering each item correctly</i>		
Choir (conditional probability)	14%	65%
Five-sided die (expected frequency)	16%	63%
Loaded die (probability ratios)	17%	55%
Mushroom (Bayesian updating)	14%	58%
<i>Gambler’s fallacy: after five heads, the next flip is...</i>		
Equally likely either way (correct)	57%	93%
More likely to land tails	24%	3%
More likely to land heads	19%	3%
Unweighted base	1,005	325

Table 4: Numeracy and the gambler’s fallacy, general public against top-PnL traders. Public: TrueDot, 21–23 September 2026, weighted to national demographics and 2024 vote. Traders: derived from TrueDot survey.

The general public scores a mean of 0.62: nearly two thirds answer no item correctly, and 2 percent answer all four. Among top-PnL traders, 13 percent score zero and 31 percent score four, so a perfect score is roughly fifteen times as common in the cohort as in the country. The gap is not driven by any single item. Traders are between 3.2 and 4.6 times as likely as the public to answer each item correctly, including the choir and mushroom items, which require conditional and Bayesian reasoning, the operations most directly involved in pricing an event contract.

A separate item asked the classic gambler’s fallacy question, the chance that a fair coin lands heads after five consecutive heads. Ninety-three percent correctly answered that the outcomes are equally likely. Among perfect-numeracy respondents the figure is 100 percent; among zero-scorers it is 74 percent, with 19 percent giving the canonical fallacious answer that tails is now more likely. The general public, asked the identical question, answers correctly 57 percent of the time, with 24 percent choosing tails and 19 percent choosing heads (Table 4). Even the least numerate traders therefore resist the fallacy more often than the public as a whole, although with a base of 43 the 17-point difference exceeds its approximate margin of ± 13 points only modestly.

6.2 Numeracy and Research Behaviour

Numeracy is not merely a test score in this population; it tracks what traders actually do when money is at stake.

When money is at stake, more likely than usual to...	0 correct	1	2	3	4 correct
Search for primary documents (filings, bill text)	14%	36%	42%	47%	54%
Seek additional and conflicting sources	47%	47%	64%	68%	71%
Use AI tools to summarise information	28%	36%	47%	43%	48%
Read full articles on what they are trading	42%	49%	43%	51%	48%
Check finance accounts on X or similar	21%	36%	26%	29%	43%
None of the above	19%	9%	13%	13%	8%

Table 5: Research behaviour by numeracy score. Stratum bases as in Table 3.

Primary-document search rises from 14 to 54 percent across the numeracy range, a spread of 40 points that comfortably exceeds the combined error on the two extreme cells. Seeking conflicting sources, the behaviour most directly associated with resistance to confirmation bias, rises from 47 to 71 percent. The share reporting no change in behaviour falls from 19 to 8 percent. Overall, 62 percent of this cohort seek conflicting sources and 42 percent search primary documents when trading.

6.3 Probability Estimates

All respondents gave probability estimates on six events resolving between late October and mid-November 2026. The relationship between numeracy and the estimates themselves is visible most clearly on the item the cohort placed furthest from even odds.

Republicans keep the U.S. House	0 correct	1	2	3	4 correct
Mean estimate	39	38	30	31	25
Median estimate	44	39	33	30	20
Share answering in the 50–59 band	26%	13%	13%	14%	11%

Table 6: Estimated probability that Republicans retain the House, by numeracy score. Stratum bases as in Table 3.

The median estimate falls from 44 among zero-scorers to 20 among perfect scorers, a difference significance tests confirm, and answers clustered around 50 falls from 26 to 11 percent. Low-numeracy traders regress toward 50 on a question the cohort as a whole places far from 50; high-numeracy traders do not. This is the same phenomenon that produces the favourite-longshot bias in prices, observed here on the demand side and indexed by a validated cognitive measure rather than inferred from behaviour.

The general-public sample answered the same six items, and Table 7 compares the two groups. On the House item the public sits at even odds. Its median is 53 against the cohort’s 30; 60 percent of the public put the chance at 50 or above, against 21 percent of traders; and 21 percent of the public answer in the 50–59 band, against 14 percent. Read alongside Table 6, the public, with a mean numeracy score of 0.62, extends the gradient below the bottom of the trader distribution: the median estimate is 53 among the public, 44 among zero-scoring traders, and 20 among perfect-scoring traders. Two qualifications apply. The House item has not resolved at the time of writing, so this is a statement about distance from even odds and the direction of disagreement, not yet about

accuracy. And public estimates on the congressional items may partly reflect partisan preference, which is known to shade election expectations, though the cohort in question appears largely politically unaligned (65 percent do not report a party affiliation) Granberg and Brent (1983).

Out of 100, the chance that...	Median estimate		In 50–59 band		Near-certain (0–9 or 90–100)		Somewhat or very confident	
	Public	Traders	Public	Traders	Public	Traders	Public	Traders
Republicans keep the U.S. House	53	30	21%	14%	13%	15%	44%	50%
Republicans keep the U.S. Senate	53	47	21%	27%	12%	3%	47%	46%
Gas above \$4.25/gal on Election Day	68	60	15%	18%	22%	8%	59%	30%
Fed leaves rates unchanged on 28 October	50	50	19%	21%	17%	8%	45%	38%
October unemployment above 4.2%	55	50	18%	28%	15%	2%	48%	22%
Harry Styles nominated for Album of the Year	50	50	21%	31%	15%	8%	39%	12%
Average across the six items	–	–	19%	23%	16%	7%	47%	33%

Table 7: Probability estimates and confidence, general public (base 1,005, weighted) against top-PnL traders (base 325, unweighted; see Section 3.5). Wording identical across surveys.

Two further patterns run across all six items. First, the public’s use of the 50–59 band is nearly constant, between 15 and 21 percent on every item. The cohort’s ranges from 14 percent on the House to 31 percent on the Grammy nomination, and outside the Senate item, which traders genuinely place close to even (median 47), it is highest where traders report the least confidence: the Grammy nomination (31 percent in the band, 12 percent confident) and the unemployment print (28 percent, 22 percent confident). Following Fischhoff and Bruine de Bruin (1999), an answer near 50 often signals that the respondent does not know. Traders appear to send that signal selectively, on the questions where it is warranted; the public sends it at much the same rate everywhere. Second, the public is more than twice as likely to give a near-certain answer, 16 percent on average against 7 percent. Fifteen percent of the public give a near-certain answer on the Grammy nomination, a question on which little information was available at fieldwork. The only item on which traders give more near-certain answers than the public is the House, and there those answers fall almost entirely on one side (15 percent at 0–9, 1 percent at 90–100), whereas the public’s are split in both directions (5 and 8 percent).

6.4 Confidence Discrimination

The six items differ substantially in how knowable they are. Respondents rated their confidence in each estimate, which permits a direct test of whether they recognise the difference. They do, and the recognition scales sharply with numeracy.

Share somewhat or very confident	General public	All traders	Top-PnL traders by numeracy score				
			0 correct	1	2	3	4 correct
Republicans keep the House	44%	50%	30%	44%	58%	46%	60%
Republicans keep the Senate	47%	46%	35%	38%	51%	43%	55%
Grammy Album of the Year nomination	39%	12%	14%	16%	17%	17%	4%
Discrimination D_s	+6.5	+36	+19	+25	+38	+28	+53

Table 8: Confidence discrimination for the general public, the full trader cohort, and by numeracy score, as defined in Section 3.3. The public and all-trader columns come from the matched comparison described in Section 3.5; the by-score columns use the stratum bases in Table 3. D_s is computed from unrounded shares.

The most numerate traders are simultaneously the most confident where confidence is warranted and, uniquely, the *least* confident where it is not: only 4 percent of perfect scorers express confidence in their Grammy estimate, against 14 to 17 percent in every other stratum. Discrimination rises from +19 to +53 across the range, though not monotonically: the three-correct stratum (+28) sits below the two-correct stratum (+38). Because the measure differences confidence within respondents, it is immune to the scale-use differences that complicate the self-description items in Appendix A.

Across the full cohort, confidence tracks item difficulty in a sensible order: 50 percent confident on the House, 46 percent on the Senate, 38 percent on the October Fed decision, 30 percent on gasoline prices, 22 percent on the unemployment print, and 12 percent on the Grammy nomination.

The general public shows almost none of this discrimination. Its score is +6.5: it is about as confident in its estimate of an unannounced Grammy nomination (39 percent) as in its estimates of congressional control (44 and 47 percent). The full cohort scores +36, and even the least numerate trader stratum, at +19, discriminates about three times as sharply as the public. The difference is concentrated where confidence is not warranted. On the two congressional items the groups cannot be distinguished (44 against 50 percent on the House, 47 against 46 on the Senate), whereas on the Grammy item 39 percent of the public express confidence against 12 percent of traders, and 16 percent of the public describe themselves as very confident against 2 percent.

The pattern holds across all six items (Table 7). The public’s confidence spans 20 points, from 39 percent on the Grammy item to 59 percent on gasoline prices, and averages 47 percent. The cohort’s spans 38 points, from 12 to 50 percent, and averages 33 percent. Traders are not more confident than the public; on four of the six items they are substantially less so. What distinguishes them is where they withhold confidence, and they withhold it most on the questions about which least can be known in advance.

6.5 Metacognition

After completing the numeracy items, respondents were asked how many they thought they had answered correctly. Forty-six percent overestimated, 42 percent were accurate, and 13 percent underestimated. The apparent gradient across strata is partly mechanical, since zero-scorers cannot underestimate and perfect scorers cannot overestimate, and we do not interpret the endpoints as evidence of a Dunning–Kruger pattern.

One feature is not mechanical and is worth recording: 28 percent of respondents who answered all four items correctly believed they had answered fewer correctly. Among the most numerate

traders on the platform, more than a quarter understate their own performance on a test they have just completed.

The general-public sample answered the same self-estimate question. In aggregate both groups overestimate: the public’s mean self-estimate is 1.68 items against an actual mean of 0.62, and the cohort’s is 2.98 against 2.42. The public’s aggregate gap is roughly twice as large, but we do not read this as evidence of better self-assessment among traders, for the reason given above: with 63 percent of the public scoring zero, overestimation is the only error most public respondents can make. Measured as the gap between the believed and actual shares answering three or four items correctly, the two groups look similar (31 against 8 percent for the public; 75 against 54 percent for traders). Comparing metacognitive accuracy within score levels would require the respondent-level public data.

7 Trading Behaviour

A number of separating trading behaviours are notable among top PnL traders.

Time commitment. Profitable traders spend a mean of 16 hours per week researching, trading, or thinking about positions, with a median of 10. Thirty-five percent spend 20 or more hours weekly and 15 percent spend 40 or more. Set against the finding that 91 percent are not full-time traders, this describes a serious second occupation rather than a hobby.

Motivation. Seventy-five percent trade to make money. Thirty percent cite the challenge of being right and 27 percent that it is fun or entertaining, with only 3 percent citing status or recognition.

Entry. The modal route in is prior market experience: 34 percent were already trading other markets, 30 percent read about prediction markets in the news or a podcast, 23 percent saw them on social media, 15 percent were introduced by a friend or family member, and 14 percent arrived via sports betting.

Privacy. Eighty-four percent trade at least partly under a handle or pseudonym (71 percent exclusively, 13 percent a mix of handle and real name), and only 11 percent discuss their trading publicly on social media, against 60 percent who discuss it with close friends and 43 percent with family. The activity is social but almost entirely private, which supplies a plausible mechanism for the mismatch between public perception and measured composition: the profitable population is, by its own choice, largely invisible.

Setting. Sixty-one percent trade from a home office or desk, 40 percent on the go, and 12 percent from work. The home-office share rises from 37 percent among zero-scorers to 72 percent among perfect scorers, consistent with the research-behaviour gradient in Table 5.

8 Discussion

Three findings organise the results.

Profitability does not require a financial career. Eighty-two percent of the platform’s highest-PnL traders have never worked full time at a bank, hedge fund, private equity firm, or proprietary trading firm. This is the finding we regard as most consequential, because it distinguishes

Kalshi from every professional venue in which people earn trading profits, and because it holds across the exchange rather than in one corner of it. Alongside it, four in five went to public high school, two in five have done customer-facing service work at rates that are near-identical across sixteen market categories, two thirds are self-taught, and nine in ten have another job (not full-time traders). Whether a trader profits in Economics or in Entertainment is unrelated to how they were schooled.

Participation on Kalshi reflects the population. The exchange-wide ZIP evidence in Section 4 is administrative, covers the whole population, and is not subject to nonresponse. Two thirds of traders live in ZIP codes below \$100,000 in median household income, the modal bands straddle the national median, and the distribution shifts downward rather than upward when sports-primary traders are removed.

What distinguishes the profitable cohort is cognitive. Numeracy in this group is high in absolute terms, and it grades almost everything else: research intensity, resistance to the even-odds heap, and above all the ability to distinguish questions that reward confidence from questions that do not. These gradients are internally consistent and, at the extremes, well outside sampling error. The general-population comparison in Section 6 shows how far the cohort sits from the country on the same items: a perfect Berlin score is roughly fifteen times as common among top-PnL traders as among U.S. adults, resistance to the gambler’s fallacy is 93 percent against 57, and confidence discrimination is +36 against +6.5. The last contrast is the most informative, because it does not reward being more confident or more numerate in general. The cohort is, on average, less confident than the public, and it is distinctive in where it withholds confidence. Part of the numeracy gap reflects education and the cohort’s other demographic differences from the public, which Section 9 addresses.

Two questions remain open, and define the natural continuation of this work.

1. **Is the cognitive difference selection or acquisition?** The design cannot distinguish whether prediction markets attract people who already reason this way or whether trading teaches them to. A longitudinal design following new accounts through their first year would separate the two, and is the obvious next step.
2. **Does numeracy pay, or merely accompany profit?** Every respondent in this study was selected on realised PnL, so the design establishes that profitable traders are numerate but cannot establish that numeracy produces profit. Crossing survey responses against PnL magnitude, trade count, and a variance measure *within* the cohort would convert a description of winners into a dose-response test. All the necessary administrative data exist.

9 Limitations

A number of limitations to this study must be acknowledged in order to holistically appreciate its value.

First, concerning ZIP code use. Section 4 characterises the neighborhoods Kalshi draws from, not the incomes of its traders. Attributing a ZIP median to an individual is the ecological fallacy, and the survey evidence makes clear that the two can diverge: the top-PnL cohort reports household incomes above the medians of typical trader ZIPs. Section 4 should be read as a statement about the exchange’s geographic and community footprint, which is what it measures well. Furthermore, we compare the trader distribution to the national median household income, which the modal bands straddle, but we do not report the share of the U.S. population living in ZIP codes in each of the same seven bands. Computing that distribution directly from ACS ZCTA tables,

weighted by households, would allow a like-for-like comparison and would either strengthen or qualify the representativeness claim. We regard this as the single most valuable addition to the present analysis. Finally, census ZIP Code Tabulation Areas approximate postal geography, are delineated decennially, and do not exist for every USPS ZIP.

Second, there are a number of structural survey features that may invite question, and if run again, should be run on a more extensive sample such as to ensure a greater degree of representativeness. The 13.8 percent response rate is high for a voluntary survey of this kind but leaves 86.2 percent of the invited frame unobserved, and traders who complete a survey containing a numeracy test may differ from those who do not in exactly the dimension the study measures (Groves, 2006). Because the frame is administratively defined, benchmarking respondents against the full invited frame on observable trading characteristics is feasible and should be done. Moreover, realised profit contains luck as well as skill. Without trade counts or a variance measure, some portion of this cohort is lucky rather than skilled, and the numeracy gradients may partly reflect a distinction between the two. More work differentiating between the two is due to be released shortly. Additionally, the public and trader samples were drawn separately, fielded over overlapping but different dates (traders 17–23 September, the public 21–23 September), and treated differently: public estimates are weighted to national demographics and 2024 vote, while trader estimates are unweighted. Neither difference bears much on the numeracy and gambler’s-fallacy items, but the forecast and confidence items could be affected by news arriving between field periods. More importantly, the cohort differs from the public in education, age, and sex, and education in particular is strongly associated with numeracy. The comparison therefore measures the distance between the cohort and the country, not the share of that distance attributable to reasoning style rather than schooling. Because the TrueDot respondent-level data exist, the public sample can be reweighted to the cohort’s education, age, and sex profile, or restricted to college graduates, to estimate how much of the gap survives. We regard this as the most important robustness check for the comparative claims in Section 6. The confidence-discrimination contrast is the least exposed to this concern, since the cohort is less confident than the public overall and differs mainly in where its confidence is placed.

Further, and on a more minor note, the public-school and service-work benchmarks are not exact counterparts to the survey items. The comparisons in Table 2 are therefore indicative.

10 Conclusions

This study set out to measure something the prediction-market literature has generally assumed: who trades, and who profits. On the first question, administrative data covering Kalshi’s entire direct trading population show an exchange rooted in middle-income America, with two thirds of traders living in ZIP codes below \$100,000 in median household income and the distribution centred on the national median of \$80,734. Removing sports-primary traders moves the distribution down rather than up.

On the second, a survey of the platform’s highest-PnL traders outside Sports and Exotics finds a population that is ordinary in origin and distinctive in reasoning. Eighty-two percent have never worked full time at a bank, hedge fund, private equity firm, or proprietary trading firm, which is the central result: at any professional trading venue that figure would be close to zero, and here it is four in five. Four in five also attended public high school. Two in five have worked in customer-facing service or retail jobs, at rates that barely move across sixteen different market categories. Two thirds taught themselves what they know about markets. Nine in ten have another job.

What separates it most sharply from the general population is numeracy and the behaviour

numeracy brings with it, measured here against a nationally weighted public sample answering the same questions: a mean Berlin Numeracy score of 2.42 against the public's 0.62, with nearly a third achieving perfection against 2 percent of the public, resistance to the gambler's fallacy of 93 percent against 57 percent, research intensity that rises steeply with measured numeracy, estimates that stay away from even odds on questions the cohort judges lopsided, and a striking ability to tell apart the questions worth being confident about from the ones that are not, in a country whose adults are nearly as confident about an unannounced Grammy nomination as about control of Congress.

Read together with the price-side evidence in Kagan and Baiocchi (2026), the picture is coherent. Calibration on Kalshi is earned market by market through time and participation, and the participants earning it are not a professional financial class. They are an unusually numerate, unusually self-taught, and largely invisible slice of ordinary America. Establishing whether that numeracy is selected or acquired, and whether it causally produces profit, is the natural next step for this research agenda.

References

- Barber, B. M., & Odean, T. (2001). Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment. *Quarterly Journal of Economics*, 116(1), 261–292.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7(1), 25–47.
- Cowgill, B., & Zitzewitz, E. (2015). Corporate Prediction Markets: Evidence from Google, Ford, and Firm X. *The Review of Economic Studies*, 82(4), 1309–1341.
- Fischhoff, B., & Bruine de Bruin, W. (1999). Fifty–Fifty = 50%? *Journal of Behavioral Decision Making*, 12(2), 149–163.
- Granberg, D., & Brent, E. (1983). When Prophecy Bends: The Preference–Expectation Link in U.S. Presidential Elections, 1952–1980. *Journal of Personality and Social Psychology*, 45(3), 477–491.
- Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646–675.
- Hayek, F. A. (1945). The Use of Knowledge in Society. *American Economic Review*, 35(4), 519–530.
- Kagan, N., & Baiocchi, R. (2026). Calibration in Prediction Markets: Theory and Evidence. Kalshi Research working paper.
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General Performance on a Numeracy Scale Among Highly Educated Samples. *Medical Decision Making*, 21(1), 37–44.
- Manski, C. F. (2006). Interpreting the Predictions of Prediction Markets. *Economics Letters*, 91(3), 425–429.
- Manski, C. F., & Molinari, F. (2010). Rounding Probabilistic Data: Alternative Data-Processing Approaches. *Journal of Business & Economic Statistics*, 28(2), 219–231.
- Mellers, B., Stone, E., Atanasov, P., et al. (2015). Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *Perspectives on Psychological Science*, 10(3), 267–281.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and Decision Making. *Psychological Science*, 17(5), 407–413.
- Tetlock, P. E. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press.
- Wolfers, J., & Zitzewitz, E. (2004). Prediction Markets. *Journal of Economic Perspectives*, 18(2), 107–126.

11 Technical Appendix

Constructing the ZIP-to-income crosswalk

Each verified residential ZIP code z is matched to the ACS median household income m_z of the corresponding ZCTA, and traders are binned by m_z . Writing N for the number of direct traders and B_k for income band k , the reported shares are

$$\hat{s}_k = \frac{1}{N} \sum_{i=1}^N \mathbb{I}\{m_{z(i)} \in B_k\},$$

with unmatched ZIPs assigned to an explicit residual band rather than excluded, so that $\sum_k \hat{s}_k = 1$. Because this is a population rather than a sample, the shares carry no sampling error; their uncertainty derives entirely from the ACS estimates m_z and from ZIP-to-ZCTA matching, neither of which is captured by a standard error.

Precision of survey estimates

For a proportion $\hat{\pi}_g$ estimated on a group of size n_g , the approximate 95 percent margin of error is

$$1.96 \sqrt{\frac{\hat{\pi}_g (1 - \hat{\pi}_g)}{n_g}},$$

which is largest at $\hat{\pi}_g = 0.5$. For the full sample ($n = 325$) this gives ± 5.4 points; for the smallest numeracy stratum ($n = 43$) it gives ± 14.9 points. A pairwise contrast between the smallest and largest numeracy strata ($n = 43$ and $n = 102$) therefore requires between roughly 13 points, for proportions near 15 or 85 percent, and 18 points, for proportions near 50 percent, to clear conventional significance. column-proportion tests at the 95

Precision of public-trader contrasts

For a difference between independent proportions from the public sample ($n_P = 1,005$) and the trader sample ($n_T = 325$), the approximate 95 percent margin is

$$1.96 \sqrt{\frac{\text{deff} \cdot \hat{\pi}_P (1 - \hat{\pi}_P)}{n_P} + \frac{\hat{\pi}_T (1 - \hat{\pi}_T)}{n_T}},$$

where deff is the design effect of the public weights, which are not reported. Under simple random sampling (deff = 1) and a conservative deff = 1.5, the headline contrasts clear this margin by wide multiples: a perfect Berlin score (29 points; ± 5.1 under both), the gambler's fallacy (36 points; ± 4.1 to ± 4.7), confidence on the Grammy item (27 points; ± 4.6 to ± 5.1), and House estimates at 50 or above (39 points; ± 5.4 to ± 5.8). The House and Senate confidence differences (6 and 1 points) do not, and the text treats them as indistinguishable. The discrimination score D combines three correlated shares from the same respondents, so its standard error requires respondent-level data; it is reported descriptively.

Treatment of 50 and of round numbers

Following Fischhoff and Bruine de Bruin (1999), a response of exactly 50 on a 0–100 chance scale may encode unfamiliarity rather than a considered even-odds judgment, and following Manski and

Molinari (2010), probabilistic responses heap at multiples of ten. Both diagnostics are computed for every respondent across the six probability items and are used to interpret Table 6, where the share of responses falling in the 50–59 band is reported alongside the medians for precisely this reason. The same band shares, and the shares of near-certain responses, are reported for the general-public sample in Table 7.

Resolution of the forecast items

All six probability items resolve between 28 October and 16 November 2026: the Federal Reserve’s October meeting, the November 3 congressional elections and the associated gasoline-price item, the early-November unemployment release, and the November 16 Grammy nominations. Individual Brier scores, and a calibration curve of stated confidence against realised accuracy, can therefore be computed for every respondent shortly after fielding. We recommend that the scoring rule be fixed in advance of resolution.

12 Appendix A: Supplementary Batteries

Base: 325 top-PnL traders unless otherwise noted. Subgroups by gender other than male are suppressed throughout for insufficient base.

A1. Reaction when a prediction turns out wrong

Response	Share
I try to figure out what I missed	64%
I update how I think about similar situations	46%
I move on quickly	21%
I'm hard on myself about it	14%
I rarely notice; I don't track my predictions	2%
I usually wasn't that confident anyway	2%

Table 9: Reported reaction to a failed prediction. Multiple selections permitted.

A2. Forced-choice cognitive style

Item	First option	Second option
Analysis even when gut disagrees / gut even when analysis disagrees	66%	34%
Think about chances of outcomes / figure out what's most likely	60%	40%
Change mind quickly / want to see more first	41%	59%
Being wrong is information / bothers me more than I'd admit	61%	39%

Table 10: Forced-choice cognitive style items.

A3. Probabilistic self-description

Statement	Describes	Neutral	Does not
I think in terms of specific chances or percentages	75%	18%	7%
I'm comfortable with phrases like "60% chance"	81%	14%	6%
I think in "will happen" / "won't happen" terms	17%	18%	65%
I consider how confident a predictor should be	70%	19%	11%

Table 11: Self-description battery, five-point scale collapsed to three categories. Note that the third item is reverse-coded: rejection indicates probabilistic thinking.

A4. Self-rated prediction ability

Sixty-four percent rate themselves better or much better than most people at predicting what will happen in the world; 30 percent about the same; 6 percent worse or much worse.

A5. Education, industry, and field of study

Highest level of education completed	Share
High school degree or less	7%
Some college, no degree	8%
Associate's degree	1%
Bachelor's degree	48%
Postgraduate degree	36%

Field of study (bachelor's or higher, $n = 272$)	Share
Business, finance, or economics	36%
Engineering or computer science	26%
Math, statistics, or physics	12%
Other natural sciences	7%
Social sciences	6%
Humanities	5%
Arts	1%
Other	7%

Table 12: Educational attainment and undergraduate field of study.

Industry (employed respondents, $n = 255$)	Share
Technology or software	26%
Banking, investment management, or hedge funds	14%
Other finance (insurance, accounting, fintech)	10%
Healthcare or pharma	6%
Engineering or manufacturing	5%
Education	5%
Retail, hospitality, or service	5%
Government or military	4%
Media or entertainment	3%
Law	3%
Other	18%

Table 13: Industry of employment among respondents employed full time, part time, or self-employed. Shares do not sum to 100 because of rounding.

A6. Probability estimates, all six items

Out of 100, the chance that...	Mean	Median	Middle 50%	% confident
Republicans keep the U.S. House	31	30	14–46	50%
Republicans keep the U.S. Senate	46	47	40–52	46%
Gas above \$4.25/gal on Election Day	60	60	49–75	30%
Fed leaves rates unchanged on 28 October	50	50	30–65	38%
October unemployment above 4.2%	46	50	32–58	22%
Harry Styles nominated for Album of the Year	47	50	34–60	12%

Table 14: Probability estimates and confidence, full cohort ($n = 325$). “% confident” is the share answering somewhat or very confident.

A7. Berlin Numeracy Test items

Item	Question wording	Response instruction	in-	Correct answer
Choir	“A town has 1,000 people. 500 are in a choir, and 100 of those are men. Of the 500 people not in the choir, 300 are men. What is the chance that a randomly picked man is in the choir?”	“Enter a percentage from 0 to 100.”		25%
Five-sided die	“A five-sided die is thrown 50 times. On average, how many of those throws show an odd number (1, 3 or 5)?”	“Enter a number from 0 to 50.”		30
Loaded die	“A six-sided die is loaded so that a 6 is twice as likely as each of the other numbers. Out of 70 throws, on average how many show a 6?”	“Enter a number from 0 to 70.”		20
Mushroom	“In a forest, 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous 20% of the time; a mushroom that is not red, 5% of the time. What is the chance that a poisonous mushroom is red?”	“Enter a percentage from 0 to 100.”		50%

Table 15: Wording, response instructions, and correct answers for the four Berlin Numeracy Test items referenced in Section 6 and Table 4. Items were fielded identically to the top-PnL cohort and the general-public sample (Section 3.5).