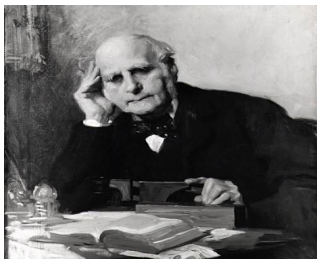
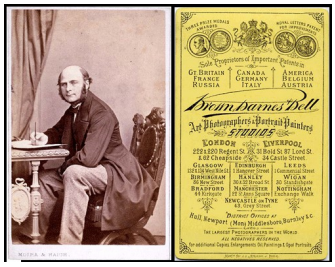
Lecture 20: Wisdom of Crowds and Prediction Markets: From Galton to Metaculus





GALTON STUDY OF WEIGHT OF OXEN: Vox Populi

In these democratic days, any investigation into the trustworthiness and peculiarities of popular judgments is of interest. The material about to be discussed refers to a small matter, but is much to the point.

A weight judging competition was carried on at the annual show of the West of England Fat Stock and Poultry Exhibition recently held at Plymouth (England). A fat ox having been selected, **competitors bought** stamped and numbered cards, for 6d. each, on which to inscribe their respective names, addresses, and estimates of what the ox would weigh after it had been slaughtered and "dressed." Those who guessed most successfully received prizes. About 800 tickets were issued, which were kindly lent me for examination after they had fulfilled their immediate purpose. These afforded excellent material. The judgements were [unbiassed] by passion and uninfluenced by oratory and the like. The sixpenny fee deterred practical joking, and the hope of a prize and the joy of competition prompted each competitor to do his best. The competitors included butchers and farmers, some of whom were highly expert in judging the weight of cattle; others were probably guided by such information as they might pick up, and by their own fancies. The average competitor was probably as well fitted for making a just estimate of the dressed weight of the ox, as an average voter is of judging the merits of most political issues on which he votes...

After weeding thirteen cards out of the collection, as being defective or illegible, there remained 787 for discussion. I arrayed them in order of magnitudes of the estimates, and converted the *cwt., quarters,* and *lbs.* in which they were made, into lbs., under which form they will be treated.

According to the democratic principle of "one vote one value," the middlemost estimate expresses the *vox populi,* every other estimate being condemned as too low or high by a majority of the voters (for full explanation see One Vote, One Value, NATURE, February 28,1907 p. 414). Now **the middlemost estimate is 1207 lb., and the weight of the dressed ox proved to be 1198 lb.; so the** *vox populi* was in this case 9 lb., or 0.8 per cent. of the whole weight too high. The distribution of the estimates about their middlemost value was ... clustered closely in its neighbourhood and became rapidly more sparse as the distance from it increased. But they were not scattered symmetrically. One quarter of them deviated more than 45 lb. above the middlemost (3.7 per cent.) and another quarter deviated more than 29 lb. below it (2.4 per cent.), therefore the range of the two middle quarters, that is, of the middlemost half, lay within those limits. It would be an equal chance that the estimate written on any card picked at random out of the collection lay within or without those limits. In other words, the "probably error" of a single observation may be reckoned as 1/2 (45 + 29) , or 37 lb. (3.1 per cent.). Taking this for the p.e. of the normal curve that is best adapted for comparison with the observed values, the results are obtained which appear in above table, and graphically in the diagram[s].

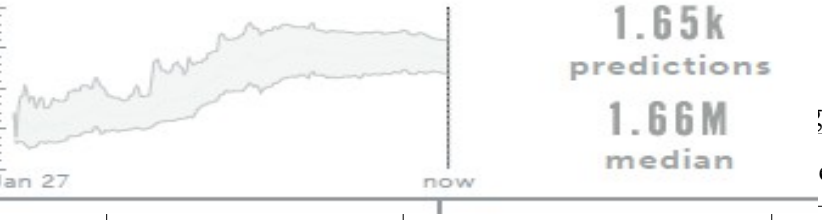
The abnormality of the distribution of the estimates now becomes manifest, and is of this kind. The competitors may be imagined to have erred *normally* in the first instance, and then to have magnified all errors that were positive. The lower half of the "observed" curve agrees for a large part of its range with a normal curve having the p.e. = 45, and the upper half with one having its p.e. = 29. **I have not sufficient knowledge of the mental methods followed by those who judge weights to offer a useful opinion as to the cause of this curious anomaly** ...

It appears then, in this particular instance, that the *vox populi* is correct to within 1 per cent. of the real value, and that the individual estimates are abnormally distributed in such a way that it is an equal chance whether one of them, selected at random, falls within or without the limits of -3.7 per cent. and +2.4 per cent. of their middlemost value. This result is, I think, more creditable to the trustworthiness of a democratic judgement than might have been expected. The authorities of the more important cattle shows might do service to statistics if they made a practice of preserving the sets of cards of this description, that they may obtain on future occasions, and loaned them under proper restrictions, as these have been, for statistical discussion. The fact of the cards being numbered makes it possible to ascertain whether any given set is complete. Francis Galton.

Metaculus.com is a community dedicated to generating accurate predictions about future real-world events by aggregating the collective wisdom, insight, and intelligence of its participants.

Users can track their predictions, earn points and powers, and hone their forecasting skills. Do you have what it takes to be a super predictor? We hope you'll join today!

Metaculus – **How many people will die as a result of the 2019 novel coronavirus (COVID-19) before 2021?**



ght of a particular living ox, made by 787 persons.

			entiles		Excess of Observed over Normal
			Observed Deviates from	Normal p.e. = 37	
	5	1074	-133	-90	+43
	10	1109	-98	-70	+28
	15	1126	-81	-57	+24
	20	1148	-59	-46	+13
q1	25	1162	-45	-37	+8
	30	1174	-33	-29	+4
	35	1181	-26	-21	+5
	40	1188	-19	-14	+5
	45	1197	-10	-7	+3
ACTUAL		**1198**			
m	**50**	**1207**	0	0	0
	55	1214	+7	+7	0
	60	1219	+12	+14	-2
	65	1225	+18	+21	-3
	70	1230	+23	+29	-6
q3	75	1236	+29	+37	-8
	80	1243	+36	+46	-10
	85	1254	+47	+57	-10
	90	1267	+52	+70	-18
	95	1293	+86	+90	-4

Galton and Metaculus are examples of "**wisdom of crowd"** in which average of views of a group of people predict actual. Better than ~95% of the individual guesses. It is a one-shot prediction market, but there are examples of a crowd that are not prediction markets. An **election opinion poll** that asks how are you going to vote is not prediction market. If it asks: who do you think will win?, it is prediction query, which beats who will vote for... "our results suggest that **expectations-based forecasts are much more powerful predictors of election outcomes**." (Rothschild and Wolfers, "Forecasting Elections: Voter Intentions versus Expectations" Jan 2013)

Gallup finds majority believe Obama will win By Jonathan Easley - 10/31/12 According to the survey, 54 percent said they thought Obama will win, 34 percent believe Romney will win, and 11 percent had no opinion. ,,, The numbers contrast with Gallup's results of who likely voters will support in the election. In its most recent poll, 51% of likely voters said they would support Romney, compared to 46% who said they would support Obama... 86% of Democrats said they believed Obama would win, compared to 71% of Republicans who said Romney. Among independents, 52% said Obama and 32% said Romney. But *The Journal of Prediction Markets (2011)* ***5*** *3, 64-74* **DO POLLS OR MARKETS FORECAST BETTER? EVIDENCE**

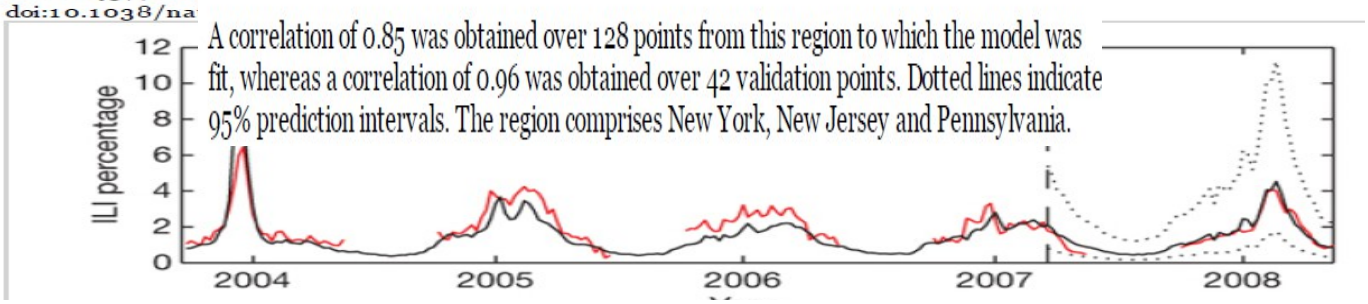
FROM THE 2010 US SENATE ELECTIONS. SAY NO DIFFERENCE. Failure in expectation equation in 2016: Sixty-eight percent of registered voters say they think Clinton will win the election, according to the poll, while 27 percent say they think Trump will win

Table 2: Forecasting the Presidential Election, by State

Year	Proportion of states where the winning candidate was correctly predicted by a majority of respondents to:		Number of states surveyed
	Expectation question	Intention question	
1952	74.3%	58.6%	35
1972	97.4%	100%	38
1976	80.3%	77.6%	38
1980	57.7%	41.0%	39
1984	86.7%	68.3%	30
1988	88.3%	56.7%	30
1992	89.4%	77.3%	33
1996	75.0%	67.5%	40
2004	89.3%	67.9%	28
2008	76.5%	76.5%	34
Totals:	279 of 345 correct	239 of 345 correct	Difference:
Average:	80.9%	69.3%	11.6%***
(Standard error)	(3.8)	(5.4)	(3.2)

Example of wisdom of crowd without market: "... analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. "

Detecting influenza epidemics using search engine query data
Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant
Nature 457, 1012-1014(19 February 2009)
doi:10.1038/na



But this turned out to be a one-shot success.

Prediction markets go beyond prediction questions. http://tippie.uiowa.edu/iem/faq.html#Real) They are markets whose assets are tied to a future event, where people can buy and sell shares based on how the future the expect. Stock markets are mechanisms that allow people to buy/shell shares in firm that should depend on expectations of future profits. Betting markets are mechanisms whose assets depend on outcomes – whether the dice comes up odd or even. Experts (bookies) set odds and adjust odds to balance predictions/bets.
Statement of economists (Science 16 May 2008, 878-888): *Prediction markets are markets for contracts that yield payments based on the outcome of an uncertain future event, such as a presidential election. Using these markets as forecasting tools could substantially improve decision making in the private and public sectors.*
Factors driving trading behavior are: risk aversion and distribution of beliefs. Prediction market prices estimate average beliefs about the probability an event occurs. The price in winner-take all prediction market is estimate of probability. Why? If market pays 50 cents for the all-or-nothing $1.00 contract that Red Sox win today, then market estimates that 1/2 of the time, will win and get $1.00 back, so this is equilibrium.
http://bpp.wharton.upenn.edu/jwolfers/Papers/Predictionmarkets.pdf; www.qmarkets.net/#resources/tutorials_pm_1; vww.ingentaconnect.com/content/ubpl/ipm)

In any area where there are views, we can AGGREGATE views via a market mechanism and generally do better than if we rely on ourselves or single experts. Why? Prediction market does the following:

1. Reflects **different information from participants** – aggregating different guesses of ox weight with market balancing biases of different people

2. **Provides feedback to participants.** By noting fluctuations in market prices, traders learn about beliefs of others and are motivated to collect and contribute more information; (but this risks herding effects).

Provides incentives to participants to reveal expertise in their trades and share knowledge anonymously

NB: Market Structure needs to be set up to rule out gaming,"information cascades", ways to manipulate outcomes.

Consider alternative **ways to cumulate knowledge– information cascade from structure**

Meet and discuss so one person can influence another . The danger is **information cascade.** This is where my assessment is influenced by yours. Say there are 2 urns: Black urn has 1/5th red balls; Red urn has ½ red balls. The experimenter selects an urn and gives **each person** a ball chosen randomly from the selected urn. The persons say the urn they think the experimenter is using – NOT the color of ball they drew

Say first person got red ball. Rational to say experimenter chose from Red uurn – it has 1/2 red > 1/5th from black urn.

Second person gets black ball. Thinks first guy must have red. One red and 1 black would most likely come from urn Red Urn which has ½ of each and so #2 says I guess it is Red urn.

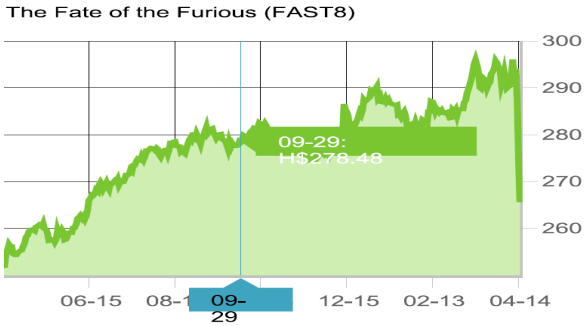
Third guy thinks, I picked black, but since the other two said Red, they likely have red balls. Best bet is Red

Fourth guy says, I have red, and others guessed red. I go with red

Fifth guy says, I have black, but if 4 people chose Red must be so I will guess with everyone else ... red

If instead each revealed what they found, would have said, we have 3 black and 2 red, best estimate is A.

Prediction Market of Movies HSX (http://www.hsx.com/), The price of a MovieStock reflects how much money Traders think the film will make For example, if The Fate of the Furious is valued at H$263.40 this means that traders expect the film to make $263.4 million in its first four weeks in wide release.

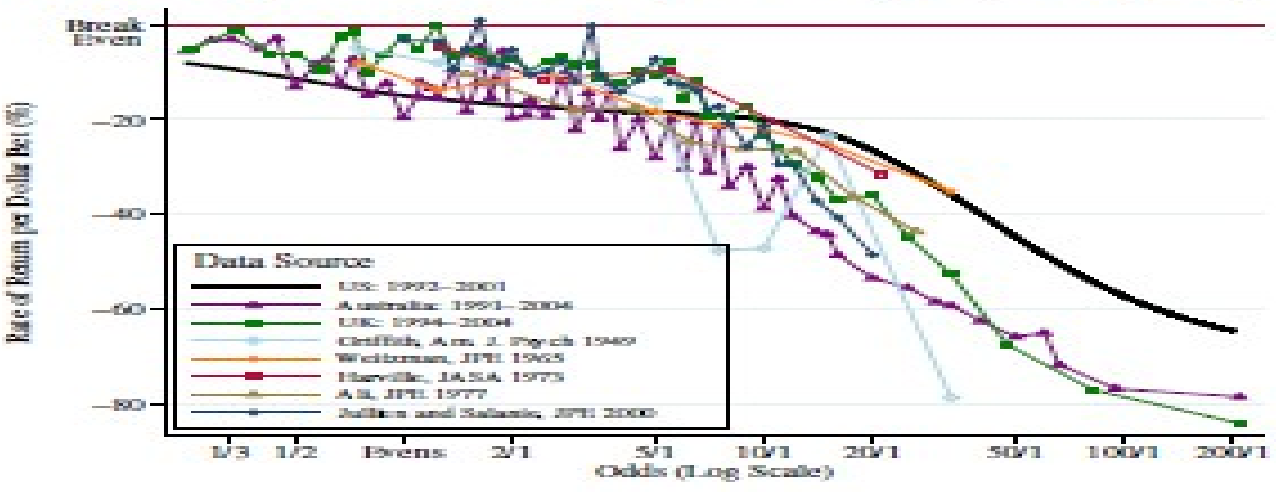


Shares Held Long on HSX: 212,088,899
Shares Held Short on HSX: 30,586,984
Trading Volume on HSX (Today):2,224,142

Other PREDICTION MARKETS: http://www.biz.uiowa.edu/iem/index.cfm; PredictIt. But many firms work with companies on internal decisions. http://www.intrade.com; https://www.cultivatelabs.com/ ; http://lumenogic.com/; https://augur.net/

Best evidence that that prediction markets works are betting markets — not perfect but very good (Sauer, JEL, Dec 1998). If the odds are 2 to 1 on a horse, then 1/3rd of time will win. If the odds are 3 to 1, then 1/4th time it will win. But there are anomalies. In perfect market where bettors seek to maximize expected value with no risk aversion no bet should have positive expected value; and the expected value of all bets should be (1- transactions cost/bet) x amount bet. But there is favorite-longshot bias. You do better betting on the favorite. (Thaler and Ziemba (Anomalies: Parimutuel Betting Markets"JEP Spring 1988)

The favorite-longshot bias is the most prominent pricing anomaly in sports betting.



Other problems: 1) Wealthy can put down more money ... lead to "wealth weighted" estimate of the event, which could differ from Vox Populi median or mean.

2) Utility function that links money to outcomes will affect bets, so risk aversion could lead people to be overly conservative. You think the odds are 50/50 but will not bet at those odds because it pains you to risk losing. Get correct odds from betting if no risk aversion and continuous distribution of views but if risk-neutral person puts all money on favorite when market price is below their view of probability -->inaccurate prediction. Manski, "interpretation of prices in actual prediction markets requires knowledge of traders' risk preferences".

3) If not real money, other preference may matter ... Justin Wolfers was top predictor on CNN presidential primary market prediction at one point. Why? He took high risk (with artificial money) to maximize chance of being top predictor ... and won. He was not revealing his best belief.

4) Traders could be non-representative of knowledgeable population.

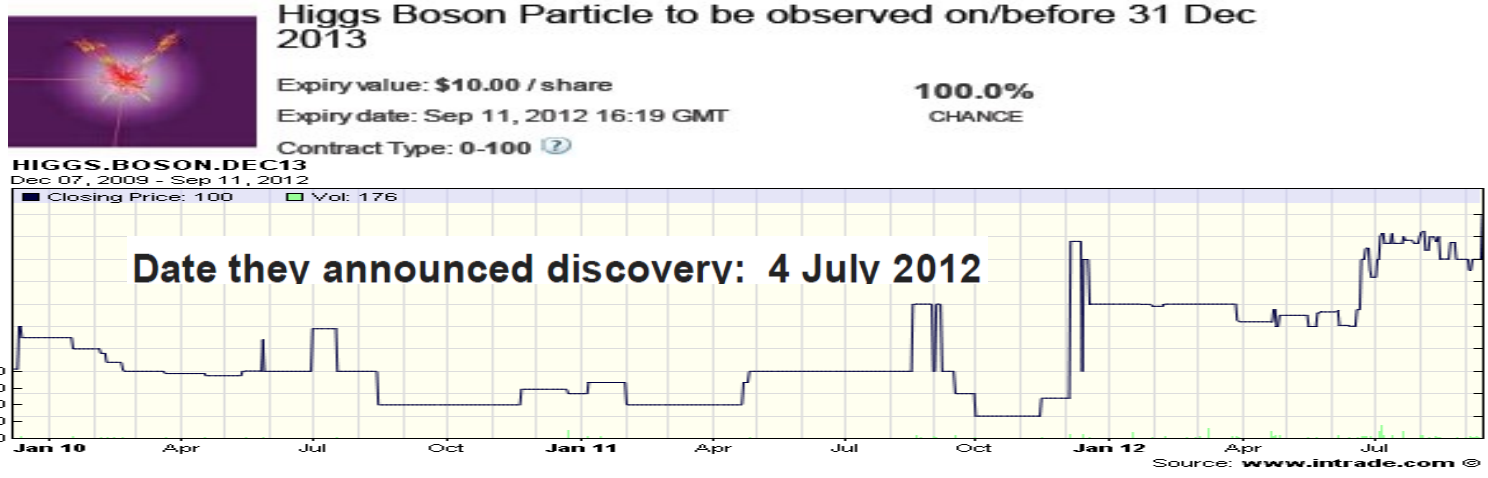
Internal Firm Markets

2003 DARPA ran betting market in future events so government could aggregate wisdom. This produced headlines that warned that terrorists could bet on attack and make money on their attack. Danger that someone manipulates events to gain in the market.

Best Buy; " We sent e-mails ...throughout the company and asked them what they thought our gift card sales would be in February 2005. We got some 190 responses and ran a simple average. It turned out to be 99.5 percent accurate, whereas the people … paid to forecast this were five percentage points off. Later that year … 350 random people predicted our holiday sales... the non-experts, off by just one tenth of 1 percent, were more accurate than the experts, ,,, off by 7 percent. These early experiments encouraged us to get into prediction contracts, and we have to date seen over 2,000 traders make a total of 70,000 trades on 147 contracts."

Google:… launched our prediction markets in April 2005, and … have asked about 275 different questions, and some 80,000 trades. Around one-quarter of our markets have to do with demand forecasting—for instance, "How many people will use Gmail in the next three months?" Another 30 percent concern the company's performance—for example, will project deadlines be met? … One important bias was optimism. **Outcomes that would be good for Google—such as getting lots of users—were slightly overpriced.** The market gave them a higher probability than it should. The cause seems to be **new employees**, whose trades show that they are highly optimistic. The external Google stock price also seems to play a role. **When Google stock does well, the price of optimistic outcomes in the prediction markets also rises.**

We also noticed that **traders under-priced extreme events**, both good and bad. When we floated contracts with five different outcomes—for example, forecasts about the number of Gmail users—the highest and the lowest outcome happened more often than the market expected... **Our markets showed that beliefs are clustered, and these clusters are made up of individuals who physically sit and work close to each other, not only at the level of city and country, but at the microlevel of the office floor, measured in feet or meters between desks.** Clusters also form around working together, socializing outside of work, and speaking a common language, even when this doesn't involve sitting close by. But these things seemed to matter less than geography. (McKinsey Report: http://faculty.haas.berkeley.edu/bo_cowgill/Popular/McKinseyQuarterlyPredictionMarketsArticle.pdf)



Wolfers and Zitzewitz (2005) conditions for prediction market prices to coincide with average beliefs of traders:

1-- Log utility based solely on expected $ return --> No consumption such as bet on underdog/ home team etc

2- Wealth and beliefs uncorrelated (if correlated instead of average belief, get wealth-weighted belief)

$$Max\ \ EU_j = q_j Log[y + x_j(1-\pi)] + (1-q_j) Log[y - x_j \pi]$$

where E is expectation; U is utility; y is initial wealth; x_j is units of prediction statement they buy; price of prediction statement on the market is Π and q is their belief

Maximize wrt X to get

$$x_j^* = y\frac{q_j - \pi}{\pi(1-\pi)}$$

The prediction market is in equilibrium when supply equal demand

$$\int_{-\infty}^{\pi} y\frac{q-\pi}{\pi(1-\pi)} f(q)dq = \int_{\pi}^{\infty} y\frac{\pi-q}{\pi(1-\pi)} f(q)dq$$

If beliefs (q) and wealth (y) are independent, then this implies:

$$\pi = \int_{-\infty}^{\infty} q f(q)dq = \overline{q}\ .$$

What if utility is not log? Works ok for plausible other utility functions.

BUT Prediction markets failed in recent elections Slate Summer 2016 and Bloomberg Nov 2016

Something's Odd About the Political Betting Markets

Brexit, Trump—the once-reliable prediction markets have misfired of late. Here's why

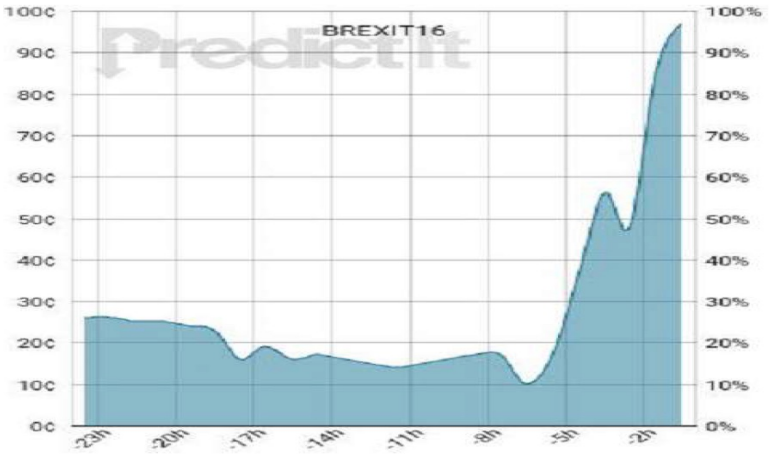
By Andrew Gelman and David Rothschild

Prediction Markets Didn't Call Trump's Win, Either

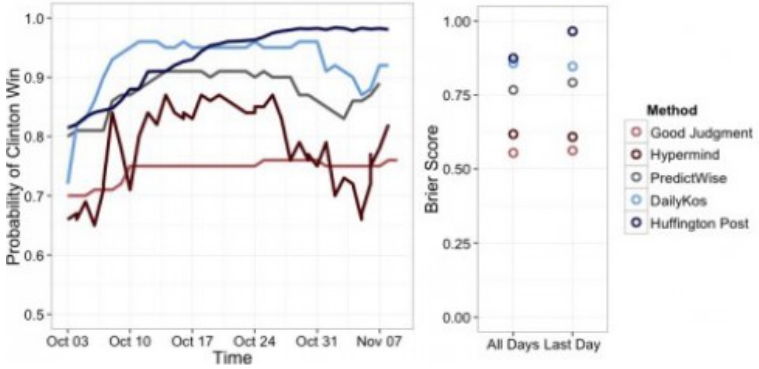
By Scott Duke Kominers

WEF, 27 Feb 2017 **Is Twitter better at predicting elections than opinion polls?**
Science, Bringing probability judgments into policy debates via forecasting tournaments

Daily price on Predictit.org for a chance to win $1 on a Trump victory



Source: Predictit.org



WHY THE PROBLEMS WITH BREXIT, TRUMP, ETC?

Interpretation 1: "prediction markets model" is flawed.

1)Gelman: "a feedback mechanism where the betting odds reify themselves" – when I go to bet in the market I note the market price and weigh that heavily in my assessment (herding; information cascade). " Traders are treating markets odds as correct probabilities and not updating enough based on outside information. Belief in

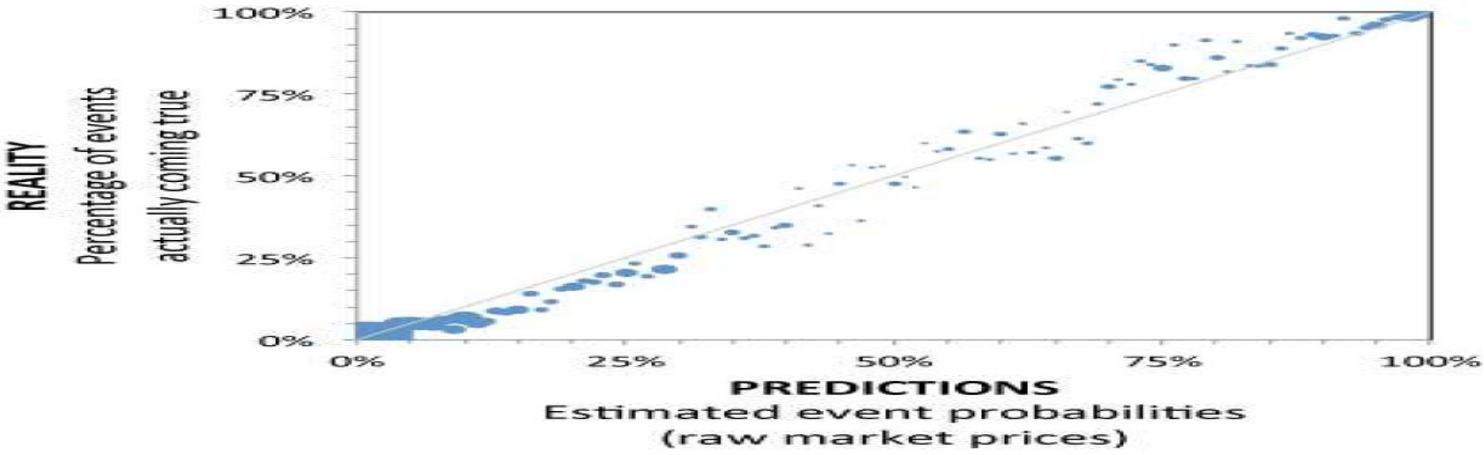
the correctness of prediction markets causes them to be too stable." (But note the jumps above!)

2)Participants are similar and see world same way. They do not add information from enough of voters not in market. Recall ensemble models supposed to include models that disagree with others and then have vote.

3)Prediction markets, people with money count more, so they would predict correctly votes weighted by $$.

Interpretation 2: Expect to fail x% of time.

Prediction markets across many events are very well calibrated.. The figure below shows close to 500 events that Hypermind prediction market predicted. If the market were perfectly calibrated, the points would fall along the diagonal line (i.e., 10% likely events would happen 10% of the time, 20% likely events would happen 20% of the time, and so on). See https://hypermind.com/hypermind/app.html?fwd=#welcome



Note pattern of errors – resembles favorite-longshot bias. You would do better in betting on high predicted events.

And Washington Post Nov 30,2016 **Which election forecast was the most accurate? Or rather: The least wrong?** Pavel Atanasov and Regina Joseph: "our analysis showed crowd-based approaches — like prediction markets and Good Judgment's prediction polls — were slightly more accurate than model-based methods.

Bloomberg https://www.bloomberg.com/politics/articles/2016-05-03/the-2016-guide-to-political-predictions-which-matter-and-who-was-most-underestimated examined 258 final projections covering 78 state primaries or caucuses from four predictors: RealClearPolitics, an aggregator of statewide polls; PredictWise, an aggregator of betting-market data; FiveThirtyEight, whose "poll-plus" prediction model considers statewide polls, national polls, and endorsements; and Bing Predicts, which combines prediction market data, polling, Internet queries, and social media posts. Of the 524 individual poll predictions collected by RealClearPolitics and HuffPost Pollster conducted within one month of a state primary or caucus, 450 of them (86 percent) correctly forecast the eventual winner. the poll aggregators included in this analysis look to be the most accurate …correctly calling nominating contests ...around 90% of the time. Prediction markets, aggregated by PredictWise, were right nearly 86% of the time.

In terms of discontinuous changes, stock markets and experts have failed to predict financial collapse, so why should we expect prediction polls to do better in politics? On Br**exit,** the financial markets were more bullish than the prediction markets in the last few days and hours before the tallies were complete.

Prediction markets succeed in predicting which psych articles could not be replicated

Using prediction markets to estimate the reproducibility of scientific research PNAS 2015

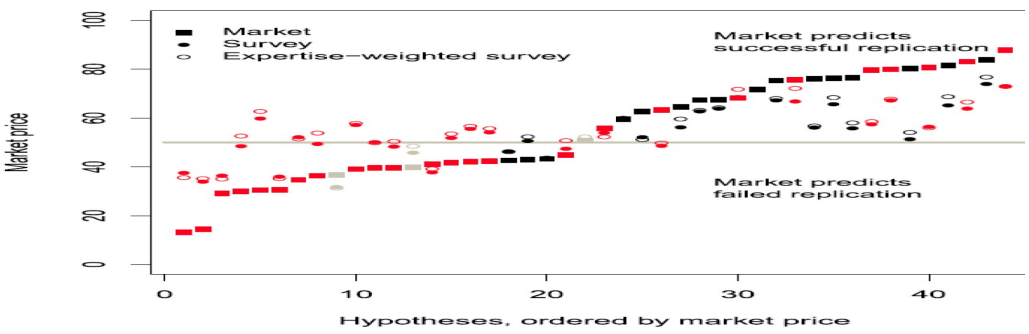


Fig. 1. Prediction market performance. Final market prices and survey predictions are shown for the replication of 44 publications from three top psychology journals. The prediction market predicts 29 out of 41 replications correctly, yielding better predictions than a survey carried out before the trading started. Successful replications (16 of 41 replications) are shown in black, and failed replications (25 of 41) are shown in red. Gray symbols are replications that remained unfinished (3 of 44).

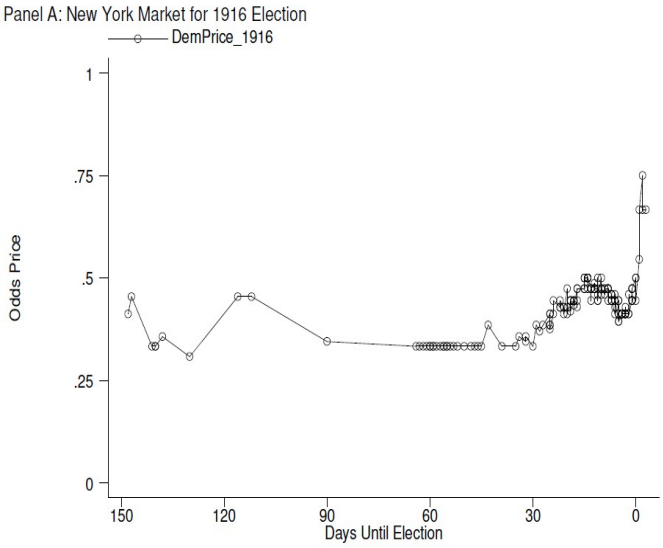
political betting markets in US : Rhode and Strumpf (2004): public and open political betting dates back to

George Washington's election. Organized election-betting markets have existed since the 1860s. From 1896 to 1924, the New York Times, Sun, and World provided price quotes almost daily... With these odds, those interested in the election could catch up quickly on its status using the aggregated beliefs of dispersed market participants to see who had the lead and by how much. Andrew Carnegie in 1904: "From what I see of the betting . . . I do not think that Mr. Roosevelt will need my vote. I am sure of his election." Erikson and Wlezien (2009) find that these markets predicted better in the era before scientific polling (election years 1880–1932) than in the era with scientific polling (1936–2008).

Table 1: New York Election Betting Volume

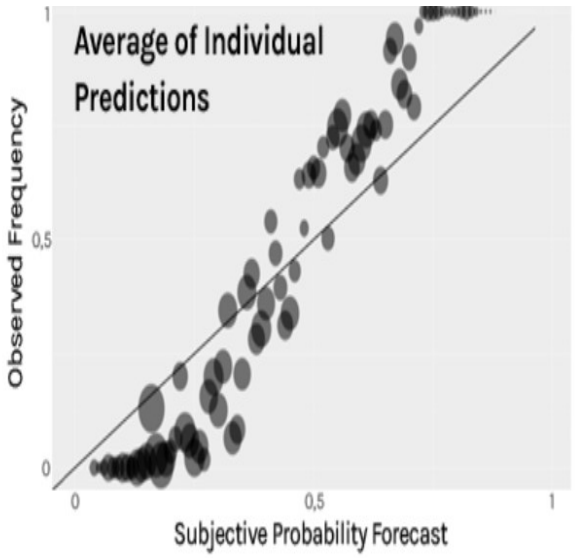
	New York Betting Volume		
	2002 dollars (millions)	Dollars per Votes Cast	Dollars per Campaign Spending
1884	13.7	1.36	0.278
1888	37.6	3.30	0.907
1892	14.8	1.23	0.185
1896	10.7	0.77	0.124
1900	63.9	4.57	0.876
1904	50.3	3.72	0.894
1908	7.7	0.52	0.174
1912	4.6	0.30	0.087
1916	165.0	8.90	2.116
1920	44.9	1.68	0.726
1924	21.0	0.72	0.373
1928	10.5	0.29	0.086
Average	37.0	2.28	0.532

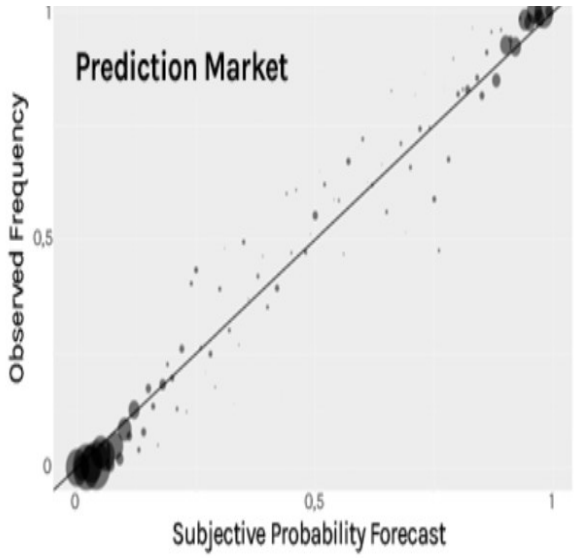
1916	Woodrow Wilson	Democratic	277	9,129,606
	Charles E. Hughes	Republican	254	8,538,221
	A. L. Benson	Socialist	0	585,113



Charts below plot the predictions of several hundred geopolitical amateurs, recruited from the general public, on a little more than 100 questions of interest to the U.S. Government. (The prediction market in this study was powered by the same Lumenogic technology that powers **Hypermind, which has prediction market on coronavirus. https://predict.hypermind.com/hypermind/app.html?fwd=#welcome .**)

The market predictions are at once much better calibrated and clearly more discriminating than those of a simple average of individual predictions.





Why it works so well

Prediction markets derive their predictive powers from three sources:

The first source is cognitive diversity. In a nutshell, when dealing with complex issues involving many variables or moving parts, no one can claim to have a complete model or theory from which to make fail-safe predictions. More likely everyone has a partial understanding of the situation, further clouded by his own biases. But when all these partial, biased models are put together, a wonderful thing happens: knowledge accumulates, gaps get filled, while the various biases cancel each other. The group's collective model is better and more complete than any individual model. Prediction markets attract diversity like powerful magnets, because anyone with a model that disagrees with the current consensus has a profit motive to participate in the market.

The second source of market accuracy is that the process encourages the voicing of informed contrarian opinions rather than calculated conformity or respectful consensus. Indeed, the only possibility for profit lies in disagreeing with the consensus, publicly. This makes sure that all informed points of views are included and aggregated. Importantly, the possibility of financial loss also discourages and penalizes the voicing of non-informed opinions.

The third source of market power is that framing predictions as wagers makes people think different, literally: Brain imaging studies show that when it contemplates a gamble, the brain becomes more risk averse and tunes out the emotional signals that might interfere with cognitive performance. In short, our thinking becomes more objective and our judgements less clouded by our passions and preferences.