

Hypermind vs. Big Data: Collective Intelligence Still Dominates Electoral Forecasting

EMILE SERVAN-SCHREIBER, Lumenogic LLC
PAVEL ATANASOV, University of Pennsylvania

Since the early 2000s, and until recently, prediction markets like *Intrade* used to dominate the election-forecasting conversation in the United States. In 2007, *The Economist* even wrote that « the most heeded futurist these days are not individuals but prediction markets, where the informed guesses of many are consolidated into hard probabilities » (Cottrell, 2007). But with *Intrade* now defunct, and the rise of sophisticated statistical models à la *FiveThirtyEight*, prediction markets attracted much less attention during the 2014 U.S. midterm election cycle. It was as if the allure of big data and statistical whizzes had eclipsed the robust and well-documented success of this form of collective human intelligence (see Servan-Schreiber, 2012 for a review). Are prediction markets doomed to become road kill on the big-data super highway?

Probably not. Previous research comparing *Intrade* to *FiveThirtyEight* during the 2008 U.S. presidential election showed that the predictions extracted from a market could outperform a sophisticated data-driven statistical approach (Rothschild, 2009). But *FiveThirtyEight* at the time was just getting started – still a work in progress – while *Intrade* was a mature, well-populated, high-profile real-money prediction market, so the comparison may not have been entirely fair, or definitive.

Our results show that the aggregated brain power of a prediction market, even a small one, even a play-money one, can outpredict a variety of the most sophisticated data-driven statistical models recently developed by the largest U.S. media organizations and academic researchers.

HYPERMIND

Hypermind.com is a new prediction market that launched in May 2014. Its trading engine is a classic continuous double auction, but liquidity is enhanced by a sophisticated friction-less arbitrage algorithm that enables bidders to match each other's orders across rival outcomes. Trades are conducted with play money – every participant starts with 100,000 H (« hyper money ») – but the best performers share real-cash prizes pro rata of their virtual profits. *Hypermind* is based on technology that *Lumenogic* developed for the *Good Judgment Project*, winner of the IARPA-funded ACE geopolitical forecasting tournament, and it incorporates some of the major insights from that groundbreaking research (Mellers et al, 2014). In particular:

(1) Good forecasting is a stable cognitive skill. It depends less on what you know (prior domain expertise) than on how you think (general intelligence, curiosity, and open-mindedness). Thus participation to *Hypermind* is by invitation only. Its several hundred participants have been actively recruited among the 20% best performers in various political, geopolitical, economic and scientific prediction-market contests run by *NewsFutures* and *Lumenogic* between 2000 and 2014. It is a small but highly-skilled group of prediction traders.

(2) Information sharing among forecasters improves overall accuracy. Thus *Hypermind* traders are encouraged to share information and help improve the collective performance by being constantly reminded that better market accuracy translates into more paying sponsors and larger cash prizes.

COMPARISON WITH FORECASTING MODELS

Hypermind listed several stocks on the U.S. 2014 midterm elections, focusing on majority-control of the Senate and the 5 most-undecided individual races in Kansas, Iowa, North Carolina, Colorado, and Georgia. This allowed comparisons between Hypermind’s predictions and those of 7 major statistical models: *FiveThirtyEight* (Nate Silver), *Washington Post*, *New York Times*, *Huffington Post*, *Princeton Election Consortium*, *PredictWise*, and *Daily Kos*. Each model had its own recipe for aggregating and weighing large numbers of polls from current and past election cycles. Some included other relevant data such as campaign fund-raising, advertising spending and macroeconomic indicators. At least one model, *PredictWise*, also incorporated predictions from the real-money U.K. prediction market *Betfair*.

The cash prize offered by *Hypermind* to be shared among its best traders on the 6 markets was just 500 euro in total (about \$650). Only 159 traders participated to the Senate-control market, conducting 1,726 trades, while the individual races markets were traded on average by just 56 traders, conducting an average of 214 trades per market.

In the analysis below we compared the predictions of each model against *Hypermind*, against each other, and against the average prediction of all the models combined. Notably, *Hypermind* predictions simply reflect the price of the last trade on each market every day (EST), without any statistical adjustments. We compared accuracy throughout the weeks or months – depending on the question – during which the market and all models were simultaneously producing predictions. The periods of comparison for each question were as follows: Senate Control [sept. 3 to nov. 4]; IA, KS, CO, NC [oct. 9 to nov. 4]; GA [oct. 20 to nov. 4].

We used Brier scores – a strictly proper scoring rule, based on squared errors. Smaller Brier scores denote higher accuracy. A perfect prediction yields a score of 0, a 50/50 prediction results in 0.5, while a perfectly inaccurate predictions earn a score of 2. To get a sense of how the methods compared overall, we computed for each question the brier score of each model every day throughout the comparison period. Then we averaged those daily brier scores into a *mean daily Brier score*. Then we averaged those across the 6 questions to get an *overall mean daily Brier score* for each model.

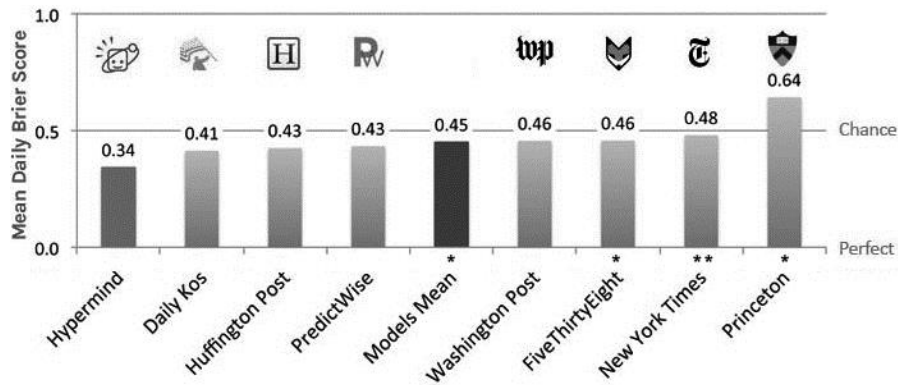


Fig. 1. Overall mean daily brier scores of *Hypermind*, the 7 statistical models, and the meta-model (Models Mean), on Senate control, and the five most undecided senatorial races: IA, KS, NC, CO and GA. One asterisk (*) indicates significant difference with *Hypermind* (paired t-test) at $p < .10$ level, while two asterisks (**) indicates significance at $p < .05$ level.

Figure 1 plots the results. All models except *Princeton’s* did slightly better than chance, but *Hypermind* out-predicted all of them, including the average prediction of all the models (“Models

Mean”). Hypermind's overall Brier score was between 17% to 46% lower than those of competing models. However, due to the small number of observations (just 6 for each model) paired Student's two-tailed t-tests reached significance only against the *New York Times* ($p < .05$), *FiveThirtyEight* ($p < .1$), *Princeton Election Consortium* ($p < .1$), and the meta-model that averages all others ($p < .1$).

We then took a closer look at these elections' most important question: would Republicans win control of the Senate? In this case, *Hypermind* again out-performed all the models, as can be seen in Figure 2 below. Except for the *Washington Post*, it appears clearly that the models remained, throughout the comparison period, much less confident than *Hypermind* in the Republican's ultimate control of the Senate. In June 2014, before the public release of all but the *New York Times* model, *Hypermind* prices stabilized around 75%, and remained in this range until October 2014.

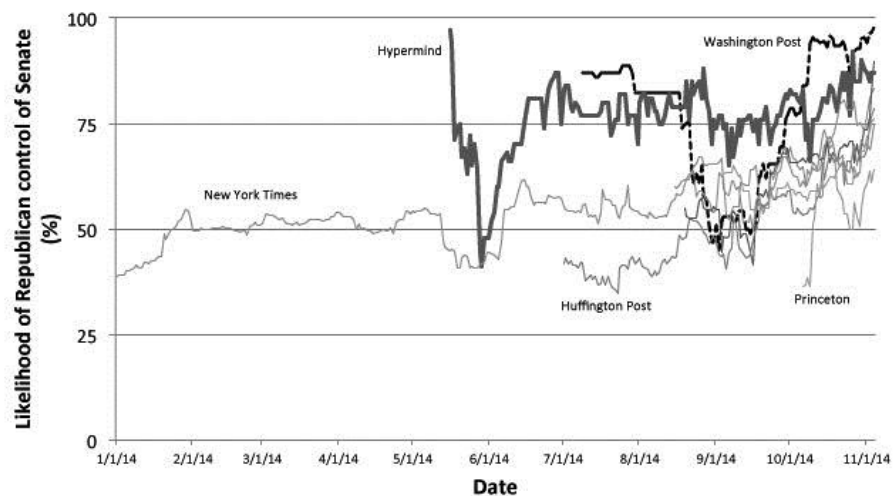


Fig. 2. Predicted likelihood of Republican Senate Control by all the models until election day (Nov. 4, 2014). *Hypermind* (heavy gray) was more confident than 6 of the 7 models throughout the race. (For clarity, *Daily Kos*, *FiveThirtyEight* and *PredictWise* curves are not individually labeled, but they appear in the bundle of curves starting in late August/early September with forecasts consistently less confident than *Hypermind*.) The *Washington Post* (heavy black dashed) is a special case that underperformed *Hypermind* in a different way (see text).

The *Washington Post* model, although highly unstable – witness the large dip around 50% in August and September – did particularly well at the end of the campaign, so *Hypermind's* advantage isn't as obvious as it is against the other models. However, when we compute the Brier scores over the period during which the *Washington Post* and *Hypermind* operated in parallel – from early July to election day – we find a 36% accuracy advantage for the market (.096) over the model (.150).

CONCLUSION

We document that a small group of skilled forecasters trading play-money predictions with the prospect of small awards was able to outperform the statistical models operated by some of the largest, most well-funded media organizations in the world. We conjecture that this can happen because, even in the age of big data, much of the information relevant to an election are still of the analog, ambiguous, implicit kind that is difficult to incorporate in statistical models. In other words, human traders incorporate data about polling, fundraising and economic outcomes, but also more subjective indicators such as debate performances and unique events. When it comes to political predictions, human collective intelligence maintains an edge over machines.

REFERENCES

- Cottrell, R. (2007). The future of futurology. *The Economist – The Word in 2008*
- Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S.E., Moore, D., Atanasov, P., Swift, S.A., Murray, T., Stone, E., Tetlock, P.E. (2014). Psychological Strategies for Winning a Geopolitical Forecasting Tournament. *Psychological Science*, 25(5), 1106.
- Rothschild, D.M. (2009). Forecasting Elections: Comparing Prediction Markets, Polls, and their Biases *Public Opinion Quarterly*. 2009. Vol. 73, No. 5, pp 895-916.
- Servan-Schreiber, E. (2012). Prediction Markets: Trading Uncertainty for Collective Wisdom, in Landemore, H. and Elster, J. (Eds) *Collective Wisdom: Principles and Mechanisms*, Cambridge University Press.