Northwestern Journal of Technology and Intellectual Property

Volume 10 | Issue 3 Article 3

Winter 2012

FantasySCOTUS: Crowdsourcing a Prediction Market for the Supreme Court

Josh Blackman
Harlan Institute

Adam Aft

Corey Carpenter

N O R T H W E S T E R N
JOURNAL OF TECHNOLOGY
AND
INTELLECTUAL PROPERTY

FantasySCOTUS
Crowdsourcing a Prediction Market for the Supreme Court

Josh Blackman, Adam Aft and Corey Carpenter

January 2012 VOL. 10, NO. 3

Volume 10, Number 3 (January 2012)
Northwestern Journal of Technology and Intellectual Property

Fantasy SCOTUS Crowdsourcing a Prediction Market for the Supreme Court

By Josh Blackman,[*] Adam Aft[**] and Corey Carpenter[***]

> The object of our study, then, is prediction, the prediction of the incidence of the public force through the instrumentality of the courts.[1]
>
> -Oliver Wendell Holmes, Jr.

> It is tough to make predictions, especially about the future.[2]
>
> -Yogi Berra

I. Introduction

¶1 Every year the Supreme Court of the United States captivates the minds and curiosity of millions of Americans—yet the inner-workings of the Court are not fully transparent. The Court, without explanation, decides only the cases it wishes. They deliberate and assign authorship in private. The Justices hear oral arguments, and without notice, issue an opinion months later. They sometimes offer enigmatic clues during oral arguments through their questions. Between arguments and the day the Court issues an opinion, the outcome of a case is essentially a mystery. Sometimes the outcome falls along predictable lines; other times the outcome is a complete surprise.

¶2 Court watchers frequently make predictions about the cases in articles, on blogs, and elsewhere. Individually, some may be right, some may be wrong. Until recently, there was not a way to pool together this collective wisdom and aggregate *ex ante* predictions for all cases pending before the United States Supreme Court.

¶3 Now there is such a tool. FantasySCOTUS.net from the Harlan Institute is the Internet's premier Supreme Court Fantasy League,[3] and the first crowdsourced prediction

---

[*] Josh Blackman is the creator of FantasySCOTUS and President of the Harlan Institute. In August of 2012, Josh will be an Assistant Professor of Law at South Texas College of Law.
[**] Law clerk for a federal judge.
[***] George Mason University School of Law, J.D. (expected May 2012). The authors would like to thank Michael Abramowicz, Ian Ayres, Tom Bell, Aaron Buhl, Miriam Cherry, Orin Kerr, Larry Ribstein, Lawrence Solum, Justin Wolfers, and Todd Zywicki. The authors dedicate this Article to F.A. Hayek and Larry Ribstein.
[1] O.W. Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 457 (1897).
[2] NASSIM NICHOLAS TALEB, THE BLACK SWAN: THE IMPACT OF THE HIGHLY IMPROBABLE 136 (2d ed. 2010). This quotation has been apocryphally attributed to Yogi Berra. *Id.* at 136 annot.
[3] With 10,000 members and rising, one writer declared FantasySCOTUS the "hottest new fantasy-league game." Bill Mears, *Frustrated with Fantasy Football? Try the Supreme Court*, CNN JUSTICE (Dec. 16,

market for jurisprudential speculation. During the October 2009 Supreme Court Term, over 5,000 members made more than 11,000 predictions for all eighty-one cases decided. Based on these data, FantasySCOTUS correctly predicted the outcome in more than fifty percent of the cases decided, and the top-ranked predictors forecasted seventy-five percent of the cases correctly. This essay explores the wisdom of the crowds in this prediction market and assesses the accuracy of FantasySCOTUS.

¶4 Part II provides an overview of the literature about the wisdom of the crowds, crowdsourcing, and prediction markets. By pooling together and aggregating the collective wisdom of many people with expansive knowledge, accurate predictions about future events can be determined to a degree of accuracy unobtainable by individuals. Part III introduces FantasySCOTUS and explains the rules of the game. FantasySCOTUS generated a novel data set with thousands of data points demonstrating how Court watchers viewed the Supreme Court and the decisions of the Justices.

¶5 Part IV assesses the accuracy of FantasySCOTUS with internal and external tests. First, to test the power of the wisdom of the crowds, this Article compares the predictions of the FantasySCOTUS "power predictors"—those who made predictions for more than seventy-five percent of the cases—with the FantasySCOTUS "crowd"—those who made predictions for less than seventy-five percent of the cases. The crowd performed worse than the power predictors, but not by much. This result lends support to the wisdom of the crowds theory, wherein a larger pool of predictors with a broader range of knowledge can often predict as well as, if not better than, so-called experts. Additionally, this Article demonstrates how the results are distinguishable from randomized results, such as coin-flips or a million monkeys playing FantasySCOTUS on iPads.

¶6 Second, this Article compares FantasySCOTUS predictions to the Supreme Court Forecasting Project's decision tree and experts, finding that the FantasySCOTUS power predictors surpassed.[4] The Forecasting Project's decision-tree relied on past voting data of the Justices to calculate the vote for any given case. The Forecasting Project also assembled a group of expert scholars and practitioners who predicted the same cases. The FantasySCOTUS power predictors predicted 64.7% of the cases correctly, surpassing the Forecasting Project's experts, though the difference was not statistically significant. The decision tree predicted 75% of the cases correctly, which is more accurate than the Forecasting Project's experts, who only predicted 59.1% of the cases correctly. The

---

2009, 2:28 PM), http://www.cnn.com/2009/CRIME/12/16/scotus.journal/index.html; *see also* Josh Blackman, *Asked About FantasySCOTUS.net in an Interview, Justice Breyer Responded: "I Don't Think I Will Bet on It,"* JOSH BLACKMAN'S BLOG (Nov. 30, 2009, 10:30 AM), http://joshblackman.com/blog/?p=2655 (quoting Justice Breyer on the importance of public interest in the Court); Bruce Carton, *Plenty of Fantasy Players on This Bench*, LEGAL BLOG WATCH (Nov. 13, 2009, 11:41 AM), http://legalblogwatch.typepad.com/legal_blog_watch/2009/11/fantasyscoutusnet-the-premier-supreme-court-fantasy-league.html (deeming FantasySCOTUS the "new gold standard in Supreme Court geekery").

[4] Theodore W. Ruger et al., *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150, 1154–55 (2004). The Forecasting Project developed a sophisticated Super Cruncher algorithm and, utilizing decision trees, predicted how the Justices during the October 2002 Term would decide cases based on certain characteristics of a case—such as circuit of origin, type of case, and the political ideology of the case. *See id.*; *see also infra* note 5 and accompanying text. To test the power of their model, the organizers of the Forecasting Project assembled a cadre of Supreme Court experts, litigators, and academics to make predictions about the same cases. Ruger et al., *supra*, at 1154–55.

FantasySCOTUS top three power predictors not only outperformed the Forecasting Project's top three experts, but also slightly outperformed the decision-tree algorithm—75.7% to 75%. This comparison provides insight into the wisdom of the crowds compared to the wisdom of specialized experts, as well as the power of a sophisticated algorithm that can "Super Crunch" the data.[5]

¶7 Part V provides an assessment of the limitations of the first version of FantasySCOTUS. As novel as these results are for the first season, FantasySCOTUS' current predictive capabilities are respectable, but not reliable—at best, it was wrong between twenty-five and thirty-five percent of the time. In light of the fact that the Supreme Court typically reverses approximately seventy percent of the cases it considers, these predictions are even less helpful.[6] FantasySCOTUS 1.0 should be understood for what it does. In its current iteration, FantasySCOTUS provides real-time *ex ante* predictions for individual cases. No other product performs this task for every case argued during the term.

¶8 Part VI considers whether a Supreme Court prediction market merely mirrors media reports about the cases—that is, whether people make predictions based on coverage about the cases in the news and blogosphere.[7] Using a comprehensive searching process—that considers both old school and new school media, such as popular legal blogs—we found a strong correlation between the amount of media attention and the accuracy of predictions. The power predictors' edge is dulled for cases that receive significant media coverage. For less popular cases that receive less media attention and about which there is less easily accessable information for prospective predictors, the crowd tends to generate less accurate predictions. In contrast, power predictors, who likely perform their own due diligence and research irrespective of media coverage, can better predict even the least noteworthy cases on the docket.

¶9 FantasySCOTUS is only two years old, but the implications and applications of this information market are intriguing. Part VII considers the possible future of FantasySCOTUS. First, from a jurisprudential perspective, FantasySCOTUS illuminates public perceptions of how the Supreme Court works as an institution. Specifically, it serves as a comprehensive polling device to provide an honest, albeit unscientific, survey that reflects how a large sample size of Court watchers view the Justices and their legal realist ideological proclivities, particularly in 5–4 decisions. If FantasySCOTUS can accurately reduce each of the Justices to nothing more than a conservative or liberal vote,

---

[5] IAN AYRES, SUPER CRUNCHERS: WHY THINKING-BY-NUMBERS IS THE NEW WAY TO BE SMART 10 (2007) ("Super Crunchers . . . analyze[] large datasets to discover empirical correlations between seemingly unrelated things. . . . Super Crunching . . . is a statistical analysis that impacts real-world decisions.").

[6] In the 2010 Term, the Court reversed seventy-two percent of merit cases before it and seventy-one percent in the 2009 Term. *Stat Pack for October Term 2010*, SUP. CT. U.S. BLOG 4 (June 28, 2011), http://sblog.s3.amazonaws.com/wp-content/uploads/2011/06/SB_OT10_stat_pack_final.pdf [hereinafter *Stat Pack OT2010*]; *SCOTUSblog Final Stats OT09—7.7.10*, SUP. CT. U.S. BLOG 10 (July 7, 2010), http://www.scotusblog.com/wp-content/uploads/2010/07/Final-Stats-OT09-0707101.pdf [hereinafter *Stat Pack OT2009*] (indicating the Supreme Court reversed 71% of its cases during the October Term 2009).

[7] Professor Orin Kerr mentioned this possibility in a 2005 blog post. Orin Kerr, *Tradesports and Supreme Court Nominations*, VOLOKH CONSPIRACY (July 19, 2005, 2:23 PM), http://volokh.com/posts/1121797428.shtml ("As a result, a site [prediction market] like TradeSports would seem to just mirror the collective common wisdom of newspapers and blogs on a question like this. Am I missing something?").

that may have broader implications to the rule of law and objective, detached standards of judging.

¶10 From a practical perspective, with more accurate future versions of FantasySCOTUS, attorneys will be able to rely on this program to assist them with litigation decisions involving cases pending before the Supreme Court. As our understanding of judicial behavior improves—perhaps through scanning all filings in PACER (Public Access to Court Electronic Records)—and the program can shift from a pure crowdsourcing technique to a commoditized Super Cruncher information service, a prediction engine can be created for lower courts. An interactive litigation assistant—think of the iPhone's Siri application—could allow attorneys and laymen alike to instantly understand and grasp the law in any given area by simply asking questions.[8] Such technology would be of great value for practicing attorneys, and provide access to justice to people who cannot afford lawyers. This is the promise of law's information revolution, of which we hope FantasySCOTUS is but a first step to the future.[9]

II. IT'S TOUGH TO MAKE PREDICTIONS, ESPECIALLY ABOUT THE FUTURE

¶11 The title of this section, apocryphally attributed to Yogi Berra,[10] recognizes the infirmity of the human mind to make predictions about the future: simply put, "we just can't predict."[11] While it is quite difficult for an individual to make predictions about the future, crowds, pooling together their collective knowledge and wisdom, are able to generate accurate predictions about unknowable events. This section explores the wisdom of the crowds, as this phenomenon is known. Prediction markets, which aggregate and assemble this wisdom, are systematic approaches to making informed predictions about the future.

A. The Wisdom of the Crowds

¶12 The wisdom of the crowds, popularized by a book by that name,[12] explores how collective knowledge can be pooled together to address problems more efficiently and accurately than decisions from individuals. The beauty of the wisdom of the crowds results from its simplicity; there are no formulas, no self-anointed experts, no normative biases from the creators of the system. Crowds are just people—people who by themselves might not be able to make consistently accurate predictions, but when pooled together generate a level of accuracy unobtainable by individuals.

¶13 "The 'wisdom of crowds' is generally more accurate and more objective than the judgment of one uninformed 'expert.'"[13] Perhaps the most popular example of the

[8] Josh Blackman, *Siri for the Law*, JOSH BLACKMAN'S BLOG (Nov. 1, 2011, 11:33 PM), http://joshblackman.com/blog/2011/11/01/siri-for-the-law/.

[9] *See* Bruce H. Kobayashi & Larry E. Ribstein, *Law's Information Revolution*, 53 ARIZ. L. REV. 1169 (2011).

[10] TALEB, *supra* note 2, at 136 annot.

[11] *Id.* at 135.

[12] JAMES SUROWIECKI, THE WISDOM OF CROWDS: WHY THE MANY ARE SMARTER THAN THE FEW AND HOW COLLECTIVE WISDOM SHAPES BUSINESS, ECONOMIES, SOCIETIES, AND NATIONS (2004).

[13] Beth Simone Noveck, *"Peer to Patent": Collective Intelligence, Open Review, and Patent Reform*, 20 HARV. J. L. & TECH. 123, 157 (2006) (quoting SUROWIECKI, *supra* note 12, at xv).

wisdom of the crowds is the "ask the audience" lifeline on the game show *Who Wants to Be a Millionaire*. If the contestant on the show is unable to answer a multiple choice question, she can pose the question to the studio audience. Instantly, the votes of each member in the audience are displayed on a screen. For over ninety percent of the questions posed to the crowd, the audience, which possesses a wider swath of knowledge, provided a correct answer where the individual contestant, who possessed a narrower range of information, could not.[14] Indeed, "[u]ncertainty is a painful part of reality; it is only natural that the wisdom of the crowd would be summoned to battle it."[15]

¶14 James Surowiecki identifies four factors to determine whether a crowd is *wise*.[16] First, the crowd must possess a diversity of opinions: "[E]ach person should have some private information, even if it's just an eccentric interpretation of the known facts."[17] Second, members of the crowd must make their decisions independently and not be influenced by others.[18] Third, all decisions should be made based only on the information available to the individual, and not based on a single, centralized source of data.[19] Fourth, the manager of the market must possess adequate algorithms to aggregate the predictions and generate accurate results.[20]

¶15 F.A. Hayek, in discussing the value of spontaneity and local knowledge, postulated that crowds, acting through markets, are better positioned to make choices than individuals who lack local knowledge.[21] To Hayek, devices such as markets are "orderly structures which are the product of the action of many men but are not the result of human design."[22] Surowiecki aimed to show that Hayek's view on the power of collective knowledge could be applied beyond descriptions of economic systems.[23]

¶16 Crowdsourcing, an application of the wisdom of the crowds, was born in a now-famous *Wired* magazine article in 2006.[24] As defined by its creator, "Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call."[25] Through crowdsourcing, "[h]obbyists, part-timers, and dabblers suddenly have a market for their efforts, as smart companies in industries as disparate as pharmaceuticals and television discover ways to tap the latent talent of the crowd."[26]

---

[14] SUROWIECKI, *supra* note 12 at 3–4.
[15] Note, *Prediction Markets and Law: A Skeptical Account*, 122 HARV. L. REV. 1217, 1217–18 (2009).
[16] SUROWIECKI, *supra* note 12, at 10.
[17] *Id.*
[18] *Id.*
[19] *Id.*
[20] *Id.*
[21] 1 F.A. HAYEK, LAW, LEGISLATION AND LIBERTY: A NEW STATEMENT OF THE LIBERAL PRINCIPLES OF JUSTICE AND POLITICAL ECONOMY, RULES AND ORDER (1973).
[22] Josh Chafetz, *It's the Aggregation, Stupid!*, 23 YALE L. & POL'Y REV. 577, 578 (2005) (quoting HAYEK, *supra* note 21, at 37) (reviewing SUROWIECKI, *supra* note 13).
[23] *Id.* at 578–79.
[24] Jeff Howe, *The Rise of Crowdsourcing*, WIRED, June 2006, at 177.
[25] Jeff Howe, *Crowdsourcing: A Definition*, CROWDSOURCING, http://crowdsourcing.com (last visited Dec. 29, 2011).
[26] Howe, *supra* note 24, at 179.

B. Prediction Markets

¶17 Building on the wisdom of the crowds, a prediction market, also known as an information market, encourages people through monetary incentives to aggregate their collective knowledge and information to predict future events.[27] People buy and sell "contracts," which effectively assign a price to the likelihood of an event happening. Several prominent prediction markets sell contracts to members that yield payments based on the outcome of an uncertain future event,[28] such as the outcome of presidential elections, returns for Hollywood movies,[29] and crime forecasting.[30] Even "data on past judicial behavior can be used to build prediction models."[31] The Iowa Electronic Markets, which pools together predictions about the Presidential election, "has yielded very accurate predictions and also outperformed large-scale polling organizations."[32]

¶18 In a prediction market, the "market price [for the contracts] will be the best predictor of the event."[33] The incentive to receive a payoff can "elicit the market's expectations of a range of different parameters."[34] Prediction markets serve three primary roles: (1) they create "incentives to seek information"; (2) provide "incentives for truthful information revelation"; and (3) generate "an algorithm for aggregating diverse opinions."[35] F.A. Hayek's writings about markets in general, where the price of goods is based on a range of information from a large group of people, accurately describe the nature of prediction markets.[36] An important value of prediction markets, beyond creating a fun forum for competitors to test their soothsaying skills, is their value as "predictive tools."[37]

---

[27] Justin Wolfers & Eric Zitzewitz, *Prediction Markets*, J. ECON. PERSP., Spring 2004, at 107, 108 ("[P]articipants trade in contracts whose payoff depends on unknown future events.").
[28] Robert W. Hahn & Robert E. Litan, *Preface* to INFORMATION MARKETS: A NEW WAY OF MAKING DECISIONS xi, xi (Robert W. Hahn & Paul C. Tetlock eds., 2006).
[29] *See* sources cited in Note, *supra* note 15, at 1218 n.6 (citing, for example, FORESIGHT EXCHANGE, http://www.ideosphere.com (last visited Dec. 4, 2011); HOLLYWOOD STOCK EXCHANGE, http://www.hsx.com (last visited Dec. 4, 2011); INTRADE, http://www.intrade.com (last visited Dec. 4, 2011); IOWA ELECTRONIC MARKETS, http://www.biz.uiowa.edu/iem (last visited De. 4, 2011), as examples of prediction markets, as well as Bill Saporito, *Place Your Bets!*, TIME, Oct. 24, 2005, at 76, for general information).
[30] M. Todd Henderson, Justin Wolfers & Eric Zitzewitz, *Predicting Crime*, 52 ARIZ. L. REV. 15, 20 (2010) ("[T]he policy-relevant question is not whether prediction markets are accurate predictors of crime rates, but whether prediction markets yield more accurate crime rate forecasts than alternative approaches.").
[31] Adam M. Samaha, *Judicial Transparency in an Age of Prediction*, 53 VILL. L. REV. 829, 834 (2008); *see also infra* note 131 and accompanying text.
[32] Wolfers & Zitzewitz, *supra* note 27, at 112; *see also* Joyce Berg et al., *Results from a Dozen Years of Election Futures Markets Research*, *in* 1 HANDBOOK OF EXPERIMENTAL ECONOMIC RESULTS 742 (Charles R. Plott & Vernon L. Smith eds., 2008).
[33] Wolfers & Zitzewitz, *supra* note 27, at 108 ("[P]articipants trade in contracts whose payoff depends on unknown future events.").
[34] *Id.* at 109.
[35] *Id.* at 125.
[36] F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 520–26 (1945) (noting that a price system permits the transfer of collective value of goods by groups, and that value could not be known by any single member of the group).
[37] Wolfers & Zitzewitz, *supra* note 27, at 112.

III. Crowdsourcing: A Supreme Court Prediction Market

¶19 FantasySCOTUS, built on the collective wisdom of its over 5,000 members, is the first and only crowdsourced prediction market for the Supreme Court. The rules for FantasySCOTUS 1.0 were simple.[38] Members could make predictions about cases argued during the October 2009 Term, up until the day the case was decided. When the Supreme Court announced in advance that opinions would be issued, voting was disabled. After an opinion was issued, all future voting for that case was disabled.[39]

¶20 Members made predictions based on eleven parameters. First, members predicted whether the Supreme Court would affirm or reverse and remand the lower court's decision. Members were awarded one point for getting the outcome correct. Second, members predicted how the Court would split: 9–0 affirm, 8–1 affirm, 7–2 affirm, 6–3 affirm, 5–4 affirm, 5–4 reverse, 6–3 reverse, 7–2 reverse, 8–1 reverse, 9–0 reverse, or other (including 4–1–4 splits or where less than nine Justices vote). Three points were awarded for correctly predicting the split. Third, for the remaining nine parameters, the members predicted whether each of the nine Justices would vote to affirm or reverse and remand. One point was awarded for each correct prediction. For a single case, members could earn up to thirteen points.

¶21 In the event that a case was not decided—for example, if certiorari was dismissed as improvidently granted—no points were awarded. Admittedly, in some cases, the scoring was difficult. In cases where a Justice voted to affirm in part and reverse in part, it was often hard to characterize whether it was an affirmance or reversal. In these cases, the rules provided that the FantasySCOTUS Czar (Josh Blackman) would isolate the most prominent part of the opinion, and determine whether a Justice voted to affirm or reverse on that issue. FantasySCOTUS 2.0 has improved rules that clarify the scoring.[40]

¶22 "The success of prediction markets, like any market, can depend on their design and implementation."[41] FantasySCOTUS is not a traditional prediction market. It is free to play,[42] contracts are not sold, and buyers are not matched to sellers.[43] At its core,

---

[38] Josh Blackman, *League Rules*, FantasySCOTUS.net, http://www.fantasyscotus.net/rules.html (last visited Dec. 4, 2011).

[39] On several occasions, due to a technical glitch, members were able to change their predictions after a case was decided but prior to the disabling of the voting feature, and effectively cheated. Those votes were eliminated, and the offenders were banned from FantasySCOTUS. Josh Blackman, *Fantasy Ethics: Cheating on FantasySCOTUS?*, Josh Blackman's Blog (Feb. 25, 2010, 12:05 AM), http://joshblackman.com/blog/?p=4198; *see also* Note, *supra* note 15, at 1222 ("Prediction markets are vulnerable to manipulation, although scholars do not agree on how serious the problem is.").

[40] *League Rules*, FantasySCOTUS, http://www.fantasyscotus.net/league-rules/ (last visited Dec. 28, 2011). The rules for FantasySCOTUS 2.0 are somewhat simpler. Rather than asking users to predict the overall outcome, and the split, the league asks users to simply predict whether a given Justice would vote to affirm. Focusing on this level of granularity allows the user to focus on each Justice and the main thrust of the case, rather than viewing the Court as a whole.

[41] Wolfers & Zitzewitz, *supra* note 27, at 120.

[42] Initially, there was no cost for law students and unemployed attorneys. Those who worked in the public sector paid a reduced fee. Those in private practice were asked to pay a nominal fee to help with site maintenance. Midway through the Term, the sign-up fee was eliminated, and everyone could play at no cost.

[43] Some research suggests that prediction markets that do not use real money may actually outperform those that force people to bet with their own wallets. Wolfers & Zitzewitz, *supra* note 27, at 120–21 ("One intriguing question is how much difference it makes whether prediction markets are run with real money or

though, it taps a "diversity of information [that] exists in a way that provides a basis for" predictions.[44] Perhaps FantasySCOTUS could be more accurately labeled a "prediction aggregation mechanism," a term coined by Professor Michael Abramowicz,[45] though for purposes of this essay, we rely on the broader conception of a prediction market.[46]

¶23 There are several potential flaws in FantasySCOTUS as a prediction market. First, some "market participants [may] trade according to their desires, rather than objective probability assessments."[47] A study suggests that participants in political markets purchase contracts in a way that reflects their party affiliation.[48] While FantasySCOTUS 1.0 did not request that members identify their ideology—we requested this information in version 2.0, and hope to elaborate on this dynamic in future work—anecdotal evidence suggests that certain members consistently voted in a manner that reflected a particular jurisprudential ideology.

¶24 In prediction markets where the "marginal trades are motivated by profits rather than partisanship, prices will reflect the assessments of (unbiased) profit motive."[49] The FantasySCOTUS market, which rewarded members with bragging rights—the grand prize was the coveted "golden gavel trophy"[50]—rather than profits, may be more susceptible to such confirmation bias. Where allowing anonymous users to make predictions without incentives may weaken the accuracy of the prediction market, it may offer the benefit of enabling a more accurate and honest polling of how the Court is perceived.[51] In other words, when not motivated by a desire to win, users may simply vote based on their personal preferences of how the Justices *should* vote—and such data are quite valuable.

¶25 Second, the outcomes of Supreme Court decisions are secret. Unlike other prediction markets where people may receive various tips about what will happen (insider trading of sorts), the votes of the Supreme Court are only known by the Justices, and their clerks (in the FantasySCOTUS rules, current clerks are banned from playing). Prediction markets "perform poorly when asked to aggregate closely guarded secret information."[52] Outside of the manner in which the question presented is phrased, discussions during oral arguments, and the questions the Justices pose, there is generally no inside information as

with some form of play money. . . . However, we do not yet have sufficient comparative data to know the extent to which money makes predictions more accurate. Indeed, it has been argued that the play money exchanges may even outperform real-money exchanges because 'wealth' can only be accumulated through a history of accurate prediction. . . . One practical advantage of play money contracts is that they offer more freedom to experiment with different kinds of contracts.").

[44] *Id.* at 120.

[45] *See* U.S. Patent No. 7,707,062 B2 (filed May 17, 2007) (discussing a "prediction aggregation mechanism").

[46] *See generally* Miriam A. Cherry & Robert L. Rogers, *Tiresias and the Justices: Using Information Markets to Predict Supreme Court Decisions*, 100 NW. U. L. REV. 1141, 1142 (2006).

[47] Wolfers & Zitzewitz, *supra* note 27, at 118.

[48] Robert Forsythe et al., *Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets*, 39 J. ECON. BEHAV. & ORG. 83, 89–93 (1999).

[49] Wolfers & Zitzewitz, *supra* note 27, at 118.

[50] *The October Term 2009 Golden Gavel Trophy*, FANTASYSCOTUS, http://www.fantasyscotus.net/news/the-october-term-2009-golden-gavel-trophy/ (last visited Dec. 28, 2011).

[51] *See infra* Part VII-A.

[52] Note, *supra* note 15, at 1225.

to how the Justices will vote.[53] This is an inherent weakness in FantasySCOTUS—and Supreme Court prediction markets generally—that could not be alleviated, short of someone with personal knowledge leaking information to the public.

¶26 Third, general criticisms of prediction markets apply equally to FantasySCOTUS. According to one critical account, "Enthusiasm for 'many minds' arguments has infected legal academia."[54] Academics "now champion the virtues of groupthink, something once thought to have only vices."[55] With respect to legal prediction markets, the criticism is somewhat more acute: "[T]he circumstances in which prediction markets are inaccurate are precisely the circumstances in which law needs them most."[56] Specifically, "the performance of prediction markets is inversely correlated with how valuable their predictions would be."[57] If an event in the future, such as the President's nominee for the Supreme Court "is secret or knowledge about its likelihood is thin, . . . a prediction market will probably not produce accurate information."[58]

¶27 Predictors "tend to overvalue small probabilities and undervalue near certainties," and "prediction markets may perform poorly at predicting small probability events."[59] "Most intractable legal informational problems involve a kind of uncertainty—whether secret, idiosyncratic, or catastrophic—not susceptible to aggregation through a market mechanism."[60] Perhaps "information markets can improve knowledge in other areas, and so indirectly improve legal decisionmaking, but this role for information markets in law is considerably more niche-like than recent scholarly enthusiasm would imply."[61] Notwithstanding these potential shortcomings, FantasySCOTUS illustrates that a legal prediction market can be accurate, reliable, and useful to the academic community, and society at large. For a number of cases, where the conventional thinking pointed to one outcome—what some may call a near certainty—FantasySCOTUS was able to discern the "small probability" vote that was generally unforeseen by experts. As FantasySCOTUS develops, these legitimate concerns about prediction markets will hopefully be assuaged.

IV. TESTING THE WISDOM OF FANTASYSCOTUS

¶28 In order to assess the predictive power of FantasySCOTUS, we devised two frameworks. The first test was internal. We compared the predictions of the FantasySCOTUS power predictors—those that made predictions for more than seventy-five percent of the cases—with the FantasySCOTUS crowd—those that made predictions for less than seventy-five percent of the cases. With this data we could not conclude that the power predictors were superior to the crowd. In other words, while we were not able

---

[53] *Id.* ("The outcomes that judges would most like to predict are naturally those about which little is already known. In the legal context, thinness of information often results from secrecy.").
[54] *Id.* at 1217.
[55] *Id.*
[56] *Id.* at 1218.
[57] *Id.*
[58] *Id.*
[59] Wolfers & Zitzewitz, *supra* note 27, at 117.
[60] Note, *supra* note 15, at 1238.
[61] *Id.*

to prove that that the crowd was just as good as the power predictors, we were able to reject the alternative hypothesis that the power predictors were simply better.

¶29 The second test was external. We compared the predictions of the FantasySCOTUS power predictors with the experts from the Supreme Court Forecasting Project, which we used as a baseline. With this approach, the diverse power predictor posse, in contrast with the largely homogenous credentialed experts from the Forecasting Project, permitted a statistically significant comparison to determine whether the wisdom of the crowds could trump the experts. The power predictors outperformed the Forecasting Project's experts, though the results were not statistically significant. Next, we compared the accuracy of the power predictors with the accuracy of the decision tree developed by the Forecasting Project. This approach allowed us to weigh the wisdom of the crowds against the power of Super Cruncher algorithms. In this case, the decision tree surpassed the accuracy of all but the best power predictors.[62]

A. Methodology

¶30 Over 5,000 members made nearly 11,000 predictions for all eighty-one cases decided during the October 2009 Term. Predictions consisted of eleven data points: the outcome of the case (affirm or reverse), the split (9–0 affirm, 8–1 affirm, 7–2 affirm, 6–3 affirm, 5–4 affirm, 5–4 reverse, 6–3 reverse, 7–2 reverse, 8–1 reverse, 9–0 reverse, and other, including 4–1–4 splits), and the votes for each of the nine Justices (affirm or reverse and remand). The analysis in this essay only focuses on the outcome of the case. Whether the Court affirmed or reversed the lower court, as opposed to the numerous splits and individual votes of the Justices, is the simplest metric to compare.

¶31 Analyzing thousands of data points required focusing on some aspect of the predictions. We decided to focus on the ten most important cases—rather than all eighty-one cases, many of which received very few predictions, and lacked statistical significance—to focus the analysis. Rather than engaging in a debate about what cases were "most important" from a normative perspective, we decided—relying on a crowdsourced approach—that the users would be the best judge of what cases mattered the most. Using the total number of predictions for each case as a measurement of popularity is particularly valuable because it does not require value judgments to determine what data matter most. The metric itself is created directly from the data set with no transformation or processing. For the purposes of the discussion, the top ten most predicted cases will be listed in descending order, from the case with the most predictions to the case with the least predictions.

¶32 First, in *Citizens United v. Federal Elections Commission*, the Court held that Congress may not prohibit corporations and unions from making independent expenditures, which are protected speech under the First Amendment.[63] It was by far the most popular case of the Term: 901 members made predictions for *Citizens United*. The second most predicted case was *United States v. Stevens*, in which the Court struck down as unconstitutional a federal statute that criminalized the depiction of animal cruelty.[64] Third was *Maryland v. Shatzer*, where the Court held that police may properly question a

---

[62] *See infra* Table 4.
[63] Citizens United v. Fed. Elections Comm'n, 130 S. Ct. 876 (2010).
[64] United States v. Stevens, 130 S. Ct. 1577 (2010).

suspect who requests a lawyer, is then released, and a couple weeks later waives his right to a lawyer.[65] Fourth was *Johnson v. United States*, in which the Court held that a prior felony does not constitute a violent felony under the Armed Career Criminal Act when the prior felony did not require the government to prove the use of force.[66] Fifth was *Padilla v. Kentucky*, where the Court found that effective assistance of counsel to an undocumented worker requires advising him or her that a guilty plea may lead to a deportation.[67] Sixth, in *Graham v. Florida*, the Court determined that a life sentence without parole for a non-capital juvenile defendant violates the Eight Amendment.[68] Seventh, in *Bloate v. United States*, the Court interpreted the Speedy Trial Act to not automatically exclude time for preparing pretrial motions; rather, the time is only excluded if the Court finds that such a delay serves justice.[69] Eighth, in *McDonald v. City of Chicago*, the Court held that the individual Second Amendment right to keep and bear arms also applies to the states.[70] Ninth, in *Perdue v. Kenny A.*, the Court found that higher than normal attorney's fees in a civil rights case are permissible only in extraordinary circumstances.[71] Tenth, in *Bilski v. Kappos*, the Court held that the Patent Act covers patentable subject matter that falls outside of the machine or transformation test.[72]

B. Defining the Power Predictors and the Crowd

¶33 To internally test the validity of the wisdom of the crowds—and whether crowds can outperform those with certain aptitudes—we focused on two groups of FantasySCOTUS members: the power predictors and the crowd. Power predictors were not selected on the basis of correctness, but rather unknowingly selected themselves by making predictions for more than seventy-five percent of the cases argued during the October 2009 Term (sixty-one out of the eighty-one cases). This group consisted of thirty members. The remaining FantasySCOTUS players—those who made predictions for less than seventy-five percent of the cases—are dubbed the crowd.

¶34 We chose not to simply pick the users with the highest accuracy rates because that would make the comparison meaningless. Picking the highest scorers would, by definition, ensure that they performed better than the crowd. Rather, we relied on the percentage of predictions as a measure of how invested users were in their predictions. Users who predicted more cases—the more popular cases as well as the obscure, less popular cases—likely devoted more effort towards predicting cases. FantasySCOTUS'

---

[65] Maryland v. Shatzer, 130 S. Ct. 1213 (2010).
[66] Johnson v. United States, 130 S. Ct. 1265 (2010).
[67] Padilla v. Kentucky, 130 S. Ct. 1473 (2010).
[68] Graham v. Florida, 130 S. Ct. 2011 (2010).
[69] Bloate v. United States, 130 S. Ct. 1345 (2010).
[70] McDonald v. City of Chicago, 130 S. Ct. 3020 (2010); *see also* Josh Blackman & Ilya Shapiro, *Keeping Pandora's Box Sealed: Privileges or Immunities, the Constitution in 2020, and Properly Extending the Right to Keep and Bear Arms to the States*, 8 GEO. J.L. & PUB. POL'Y 1, 17–22 (2010) (discussing *McDonald*); Alan Gura, Ilya Shapiro & Josh Blackman, *Extending the Right to Keep and Bear Arms: The Tell-Tale Privileges or Immunities Clause*, 2010 CATO SUP. CT. REV. 163 (2010).
[71] Perdue v. Kenny A., 130 S. Ct. 1662 (2010).
[72] Bilski v. Kappos, 130 S. Ct. 3218 (2010).

top three power predictors—those who made predictions for 75.7% of the cases—collectively fell only one prediction short of offering predictions for all of the cases.

¶35 The composition of the power predictor posse is quite varied, and members fell into five general types of members.[73] First, a number of power predictors had some Supreme Court experience, mostly with writing amicus briefs in law school Supreme Court clinics. The most accomplished player, solo practitioner David Mills,[74] successfully argued and won *Ortiz v. Jordan*.[75] Second, a few power predictors were professors—both of law and political science. One taught as an Associate Professor at Columbia Law School, another as a Political Science Professor at Rice University. Third, the vast majority of power predictors were law students and new attorneys. Most notable among this group was the champion, "Chief Justice" Justin Donoho,[76] who recently graduated from the University of Chicago, worked as a law clerk on the Seventh Circuit, and served as a research assistant to Judge Richard A. Posner.[77] Other top student power predictors attended Southern Illinois School of Law, Vermont Law School, and the University of Tulsa.

¶36 Fourth, some power predictors were attorneys who lacked appellate experience. One member in this group has a small general practice firm with his wife in Alabama. Finally, the most interesting group of power predictors consisted of players who had no formal legal training and never attended law school. One of the best players is an actuary who never attended law school and, quite impressively, taught himself constitutional law in high school. He was the "Chief Justice" of the October 2010 Term of FantasySCOTUS.[78] The eighth-ranked player never attended law school and works as an air traffic controller. The eighteenth-ranked player is not a lawyer and earned a degree in Geophysical Engineering from Montana Tech.

¶37 Admittedly, our selection of the power predictors is somewhat flawed in that they were selected *ex post*. As a practical matter, we had no other choice. There was no way to select a group of top players *ex ante* in the first season of FantasySCOTUS, when everyone started with a collective score of zero. Similarly, setting the prediction level at seventy-five percent was arbitrary, as we had no other historical baseline to rely on. Any accuracy derived from the increased participation was not deliberate on our part, although incidentally our power predictors effectively overlapped with the top-ranked players.

¶38 As our analysis suggests, we can state to a degree of statistical significance that the power predictors' results were not based on chance—that is, predicting that every case the Court hears would be reversed, for example, with hopes of guessing one's way to a high score. For the second season, however, the top performers (those with the most

---

[73] Some, but not all of the power predictors responded to a survey inquiring about who they were and how they made predictions. Some of those who responded requested various degrees of anonymity. For a detailed discussion of who they are and how they made their predictions, see Josh Blackman, *Who Are the FantasySCOTUS Experts and How Do They Make Predictions?*, JOSH BLACKMAN'S BLOG (May 4, 2011, 11:14 PM), http://joshblackman.com/blog/?p=6875.

[74] Mark Curriden, *The Long Shot*, A.B.A. J., Nov. 2010, at 52.

[75] *See* Ortiz v. Jordan, 131 S. Ct. 884 (2011).

[76] *Congratulations to Justin Donoho, the Chief Justice of FantasySCOTUS OT2009*, FANTASYSCOTUS, http://www.fantasyscotus.net/?p=94 (last visited Dec. 28, 2011).

[77] *Id.*

[78] Josh Blackman, *JoshCast: Interview with Jacob Berlove, Chief Justice of FantasySCOTUS OT 2010*, JOSH BLACKMAN'S BLOG (July 5, 2011, 9:01 AM), http://joshblackman.com/blog/2011/07/05/joshcast-interview-with-justice-donoho-chief-justice-of-fantasyscotus-ot-2010/.

accurate predictions) from the first season who returned are designated as repeat power predictors. We are interested to see how repeat performers do. In future work, we will select the next generation of power predictors.

C. Comparing the Power Predictors and the Crowd

¶39 We compared the accuracy of the crowd with the accuracy of the power predictors using two tests. First, we compared the outcome accuracy rate of the two tests—that is, whether they correctly predicted that the Court would affirm or reverse a case—as a percentage. This approach allowed us to focus on one clear metric, the outcome. Second, we considered all parameters of the prediction—the outcome, the split, and the votes to the individual Justices—and compared the total scores of the crowd and the power predictors. This approach provides a more comprehensive analysis to see how granular and detailed the accuracy of the crowd is—it is much more difficult to predict the individual votes than to simply predict an overall affirm or reverse outcome. Based on these two approaches, we could not conclude that the power predictors' predictions were superior to those of the crowd, supporting the validity of the wisdom of the crowds.

1. Accuracy of Forecasting the Outcome

¶40 Table 1 presents the outcome of each case, the crowd to power predictor ratio, the accuracy of the crowd,[79] whether it was significant,[80] the power predictor accuracy rate, whether it was significant, and the date of oral arguments.

TABLE 1. GENERAL GROUP STATISTICS

Case	**Outcome**	**Crowd-to-Power Predictor Ratio**	**Crowd**		**Power Predictor**		**Number of Predictions**	**Oral Argume nts Date**
			Accuracy	**Outcome Sig.**	**Accuracy**	**Outcome Sig.**		
Citizens United	Reverse	24.743	61%	Yes (99%)	71%	Yes (99%)	901	9/9/09
Stevens	Affirm	16.135	83%	Yes (99%)	92%	Yes (99%)	634	10/6/09
Shatzer	Reverse	15.946	50%	No	65%	Yes (90%)	627	10/5/09
Johnson	Reverse	6.868	45%	No	50%	No	299	10/6/09
Padilla	Reverse	7.027	38%	Yes (99%)	35%	Yes (90%)	297	10/13/09
McDon ald	Reverse	6.459	66%	Yes (99%)	65%	Yes (99%)	276	3/2/10

[79] The accuracy rate, expressed as a percentage, represents the percentage of each group that correctly predicted the outcome of a case.

[80] We calculated whether each of the outcomes—crowd and power predictors—was statistically significant, based on confidence intervals. In statistics, the determination of reliability—that is, how likely the data express a clear outcome (affirm or reverse)—can be specified based on various confidence intervals. The most commonly used intervals, 90%, 95%, and 99%, are sorted in increasing reliability. A 99% confidence level, the gold standard of statistical measures, indicates that the sample results are most reliable. In contrast, 95% and, even more so, 90% confidence intervals express that the data are less likely to be reliable. Each group is independently above the threshold for the Central Limit Theorem. The larger the crowd, the more accurate the results are.

Graham	Reverse	5.486	49%	No	60%	No	240	11/9/09
Bloate	Reverse	5.405	20%	Yes (99%)	32%	Yes (95%)	237	10/6/09
Perdue	Reverse	4.514	57%	Yes (90%)	62%	No	204	10/14/09
Bilski	Affirm	4.270	80%	Yes (99%)	78%	Yes (99%)	195	11/9/09

¶41 Due to the novelty of the data set, we created a custom decision rule to determine if each group, as a whole, was correct (or incorrect) above a certain threshold. A confidence interval is a range where the values of the test statistics—in FantasySCOTUS, the affirm or reverse percentage—may differ due to statistical noise.[81] Statistical noise could manifest as uncertainty introduced through sampling.[82] A determination was reliable when the confidence interval did not include 50%, meaning that enough predictions for either affirm or reverse were—for statistical purposes—in agreement with the direction of the outcome such that we could reliably assess the prediction at a given confidence interval. This analysis yields a definitive affirm or reverse decision for the group. Otherwise, the predictions would not be conclusive for affirmation or reversal.

¶42 The data suggest several statistically significant trends. First, the number of predictions made roughly tracked the date of oral arguments. The earlier the case was argued, the more predictions were made. Generally, members made predictions following oral arguments. On November 11, 2009 when FantasySCOTUS launched and went viral, *Citizens United*, *Stevens*, and *Shatzer*, the three cases with the highest crowd-power predictor ratio—which received 901, 634, and 627 predictions, respectively—had already been argued. However, *McDonald*, a landmark Second Amendment case not argued until March 2, 2010—nearly four months after FantasySCOTUS launched—received only 277 predictions. It seems that the number of predictions dropped off over the course of the Supreme Court Term, and web traffic analytics anecdotally support this conclusion.

¶43 Cases argued later in the Term had fewer members of the crowd make predictions. Yet the dedicated detail of power predictors voted consistently throughout the Term. Additionally, the accuracy of the predictions seems to indicate that the weakest predictors of the crowd left at an early stage. In this sense, cases argued later in the Term had a more reliable set of predictors.

¶44 Second, there was an interesting interplay between the statistical significance of the data and the accuracy of the resulting predictions. Statistical significance must be distinguished from accuracy. Significance refers to how reliable the data are, while accuracy refers to how correct the data are—the difference between precision and correctness.[83] For all cases, in light of the smaller sample size, the power predictor group had a larger confidence interval—meaning more statistical noise and less reliable results. Yet, the power predictor group still provided more accurate results. Conversely, the

[81] B.S. EVERITT, THE CAMBRIDGE DICTIONARY OF STATISTICS 86 (2d ed. rev. 2003) ("Confidence interval: A range of values, calculated from the sample observations, that are believed, with a particular probability, to contain the true parameter value.").
[82] *Id.* at 332 ("Sampling error: The difference between the sample result and the population characteristic being estimated. In practice, the sampling error can rarely be determined because the population characteristic is not usually known.").
[83] JOINT COMM. FOR GUIDES IN METROLOGY, INTERNATIONAL VOCABULARY OF METROLOGY—BASIC AND GENERAL CONCEPTS AND ASSOCIATED TERMS 21–22 (2008).

crowd had a smaller confidence interval—less statistical noise and more reliable results—but produced less accurate results. This illustrates the difference between reliability viz. statistical significance and accuracy. The crowd was more reliably wrong, meaning that, while the results were a reliable expression of the crowd's predictions, those predictions were inaccurate. In contrast, the power predictor group was less reliable but more accurate.

¶45 The predictions for eight of the ten cases, excluding *Johnson* and *Graham*, were statistically significant. In this context, statistical significance means that the data show a definitive affirm or reverse outcome according to our decision rule. For *Johnson* and *Graham*, which were not statistically significant, the prediction data for affirm and reverse are statistically 50/50—equally likely to generate an affirm or reverse result—and thus were inconclusive. The prediction data for *Padilla* and *Bloate* provide for statistically significant results, yet the ultimate affirm-or-reverse predictions were incorrect. For the remaining six cases, which were decided correctly, the prediction data provide for statistically significant results.

¶46 For *Shatzer*, the power predictor group's predictions proved statistically significant, while the crowd's predictions were not statistically significant. This split represents an exception to the rule, because generally, a smaller sample size (power predictors) would not be statistically significant, whereas a larger sample size (crowd) would be statistically significant. Here, the power predictor group (thirty members) was more statistically significant than the crowd (about 590 members). This suggests that, for *Shatzer*, the power predictor group was more reliable—that is, a decisive consensus existed with respect to the affirm or reverse outcome.

2. Accuracy of Forecasting Outcome, Split, and Individual Votes

¶47 Members of FantasySCOTUS made predictions for the outcome, split, and individual votes of each Justice. A perfect score was thirteen.[84] Table 2 calculates two averages for each case—the average score for each member of the power predictor group (Power Predictor Average Points) and the average score each member of the crowd (Crowd Average Points). For each average, we calculated the standard deviation.[85] As an additional measure of statistical difference between the two groups, we calculated the Welch's *t*-test.[86] To measure overall performance (that is, how each group performed

[84] *See supra* Part III.

[85] The standard deviation for each average reveals how individual scores—that is, the votes of individual members in each group—were distributed around the average. Standard deviation is a common statistic used to measure the spread of data points within a sample. EVERITT, *supra* note 81, at 360. The standard deviation is equal to the square root of the variance. *Id.* For example, if the average score was nine and the standard deviation was one, that would indicate that a certain number of members, within one standard deviation, scored between eight and ten points. A larger standard deviation indicates that there was a greater spread from the average for members in that group as a total.

[86] If two data points are significantly different, that suggests that the difference is not due to statistical noise or randomness. A Welch's *t*-test, using standard deviation, determines whether the two averages were significantly different from each other. We indicated if the results were statistically significant, and whether the confidence level was 90%, 95%, or 99%. *See generally* B.L. Welch, *The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved*, 34 BIOMETRIKA 28 (1947). A *t*-test is a commonly used statistical test to determine if two samples are significantly different through hypothesis testing. A Welch's *t*-test is appropriate in this case, due to its relaxed assumptions, and

over all the cases), we averaged together each individual average from the ten cases—in other words, it is an average of the averages. Based on this average, we calculated the overall standard deviation for each classification.

TABLE 2. AVERAGE SCORES

Case	Crowd		Power Predictor		Significant? (Welch's *t*-test)
	Average Points	Standard Deviation	Average Points	Standard Deviation	
Citizens United	8.62	4.16	9	4.22	No
Stevens	8.17	2.53	8.57	2.23	No
Shatzer	6.9	1.86	8.11	2.55	Yes (99%)
Johnson	6.18	2.41	7.08	2.97	Yes (90%)
Padilla	5.65	2.48	6.08	2.61	No
McDonald	9.17	3.5	9.68	2.99	No
Graham	5.42	3.07	6.49	2.91	Yes (95%)
Bloate	6.56	1.38	6.81	1.58	No
Perdue	6.65	2.86	6.75	2.86	No
Bilski	9.22	3.34	10.68	3.1	Yes (95%)
Avg. of Avg.	7.25	1.43	7.93	1.53	No

¶48 For all ten cases, the average member of the power predictor group scored more points than the average member of the crowd. The power predictor average, 7.93 points, was higher than the crowd average, 7.25 points. The biggest difference was in *Bilski*, where the power predictor group scored on average 10.68 points, while the crowd group on average scored only 9.22 points, a difference of 1.46. The case with the least difference was *Perdue*, where the power predictor group scored on average 6.75 points, while the crowd group on average scored only 6.65 points, a difference of 0.1.

¶49 While members of the power predictor group scored more points on average, they generally had a higher standard deviation—in other words, a larger point spread—around that average than members of the crowd group. Using *Shatzer* as a typical example, the standard deviation of the power predictor group was 2.55, more than 33% greater than the 1.86 standard deviation of the crowd group. This indicates some uncertainty within the power predictors' forecasts. Perhaps power predictors, convinced of their individual views of the case, are more likely to buck trends and make less conventional predictions about individual Justice behavior. The crowd tends to be more unified in this sense; adhering to conventional views of how the Justices will vote perhaps indicates that crowds are more influenced by media coverage.[87]

does not require that the variance within a sample be the same. Without the data set, we could not ensure that the variance would be the same between different samples.

[87] *See infra* Part VI-A.

¶50 The Welch's *t*-test in six of the ten cases yielded no significant results. That is, in six of the ten cases the crowd was just as likely as the power predictors to predict the correct outcome. Further, in light of the extensive media coverage of *Citizens United*, *Stevens*, and *McDonald*,[88] any informational advantage the power predictors might have had over the crowd was minimized. For *Shatzer*, *Johnson*, and *Graham*, the difference between the two groups was significant. In these cases, the power predictors may have had insight that the crowd did not—perhaps they gleaned some clue from oral arguments that the media overlooked or discerned how the Justices would vote from the arguments in the briefs.

3. Analysis

¶51 The results do not conclusively prove that the power predictors' forecasts were superior to those of the crowd. Although the power predictors generally do better, the crowd is able to make rather strong predictions to bridge the gap. This lends support to the wisdom of the crowds theory, wherein a larger pool of predictors with a broader range of knowledge can often predict as well, if not better, than knowledgeable individuals.

¶52 However, this does not hold true for all cases. In the marginal cases, the crowd performs well, but just not as well as the power predictors do. Generally, the power predictors make more informed predictions, although the predictions lack high levels of accuracy.[89] The lack of precision could very well reflect a professional hubris of sorts. With too much knowledge, and perhaps over-confidence, the power predictors may have second-guessed conventional wisdom, and prudence. In contrast, the crowd is more unified in its results, and perhaps influenced by extensive media coverage of the cases.[90] In this respect, the power predictors exhibit some of the flaws particular to experts, and these results demonstrate how a crowd can smooth out these errors.[91] In summary, the power predictors are only better predictors in the marginal case. FantasySCOTUS' wise crowds are about as accurate as the power predictors, meaning individuals who only make a few predictions, when aggregated, were almost as accurate as those who made many predictions.

4. Insulating Results from Randomness

¶53 Our statistical modeling, combined with the disparate nature of the Supreme Court docket, helps to mitigate the risk of randomness weakening the reliability of the comparisons and trends. The outcome of one case, generally, will not affect the outcome of a second case (unless they are precedentially related). For example, *United States v. Stevens*, decided early in the Term, had no impact on *Bilski v. Kappos*, decided at the end of the Term. These cases are *independent trials*—the voting in past cases has no impact

---

[88] *See infra* Part VI-A.
[89] Blackman, *supra* note 73 (discussing the sources power predictors rely on to make their decisions).
[90] *Id.*
[91] MICHAEL ABRAMOWICZ, PREDICTOCRACY: MARKET MECHANISMS FOR PUBLIC AND PRIVATE DECISION MAKING 38 (2007) (Aggregated data may "in part cancel out random errors that individuals make in predictions by overweighing or underweighing particular pieces of evidence.").

on the voting in future unrelated cases.[92] Or, to put it another way, flipping a coin once has no impact on flipping the coin a second time.[93]

¶54 When comparing a user's performance to a possible random performance—flipping a coin for every prediction—we consider the cumulative results of individual votes, rather than the overall score.[94] In other words, we count the number of times a coin landed on the correct side, rather than the total number of heads or tails. Because the cases are independent due to the different legal doctrines and factual predicates, it is much harder for a user to randomly make predictions to obtain a high score over the course of the entire Supreme Court Term with eighty-one cases decided by the same set of Justices. Specifically, where we presented data we indicated whether they were significant and at what confidence level (90%, 95%, or 99%). With data at these confidence levels, we were able to assert that the result was not due to randomness or chance.

¶55 Our methodologies also prevent a statistical "quasi-miracle,"[95] whereby users could randomly predict all of the cases accurately. Assume an infinite number of monkeys were stationed at FantasySCOTUS iPads, randomly making predictions,[96] and obtained a perfect score. A single primate, let's call him Ape Fortas, would need to correctly predict each and every case. The odds of Ape Fortas accomplishing this task are infinitesimally small.[97] Even 5,000 (the number of FantasySCOTUS players) apes mashing away on five-thousand monkey-friendly iPads would not increase the odds of any one player predicting all of the cases correctly. This small sample size is not even close to the same

[92] STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 477–79 (2d ed. 2003) (discussing the various forms of independence used in decision making and probability theory).

[93] Formally, coin flips belong to a class of stochastic processes called Bernoulli processes, which use a constant probability of binary outcomes in multiple trials. CHARLES M. GRINSTEAD & J. LAURIE SNELL, INTRODUCTION TO PROBABILITY 96–97 (2d rev. ed. 1998), *available at* http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf.

[94] A binomial distribution is the creation of a probability distribution depending on constant probability, the number of trials, and the number of successes to determine the likelihood of conditional outcomes. Michelle Lacey, *The Binomial Distribution*, YALE U. DEPARTMENT STAT., http://www.stat.yale.edu/Courses/1997-98/101/binom.htm (last visited Dec. 28, 2011).

[95] Quasi-miracles are thought of as the logical equivalent of denying the existent of absolutes, i.e. objects always fall towards the ground or a series of a million coin flips will come up heads. For practical purposes, such events are extremely rare, but are an important part of logical statements. *See* J. Robert G. Williams, *Chances, Counterfactuals, and Similarity*, 77 PHIL. & PHENOMENOLOGICAL RES. 385, 389 (2008).

[96] The earliest mention of the monkey thought experiment, where one of an infinite number of monkeys, pounding away at typewriters, produces the complete works of Shakespeare, was introduced by Émile Borel, a French mathematician in 1913. Émile Borel, *La Mécanique Statique et L'irréversibilité* [*Static Mechanics and Irriversibility*], 3 JOURNAL DE PHYSIQUE THÉORIQUE ET APPLIQUÉE [J. PHYS.] 189 (1913) (Fr.).

[97] Richard Dawkins lays out the probability of writing a twenty-eight character sentence from Shakespeare by a monkey on a typewriter with just the twenty-six letters and the spacebar as $(1/27)^{28}$. RICHARD DAWKINS, THE BLIND WATCHMAKER: WHY THE EVIDENCE OF EVOLUTION REVEALS A UNIVERSE WITHOUT DESIGN 46–47 (3d ed. 1996). The monkey has a 1 in 27 chance of getting each character right, but must get the characters correct in sequence, causing the exponential probability. *Id.* at 47. In our case, Ape Fortas would face the odds of $(1/2)^{61}$ (sixty-one represents 75% of the eighty-one cases decided). Although this number is much larger than the monkey's odds, it would still be highly unlikely to occur.

order of magnitude to compare with all those prescient prognosticating primates.[98] The approach we used attempted to insulate the data from logical and statistical problems of randomness. Further, at the conclusion of the Term we verified that no member of FantasySCOTUS made predictions by selecting reverse for every case, thus confirming that predictions were not made randomly, at least using this strategy.

D. The Supreme Court Forecasting Project

¶56 In a path-breaking article, a group of legal and political science "Super Crunchers"[99] developed a decision tree model, based on the prior decisions of the nine Justices, to predict outcomes of Supreme Court cases during the October 2002 Term.[100] The article compared those outcomes to the predictions of a group of experts.[101] The Project's decision tree did not consider the legal merits of a particular case. Instead the authors based their model on six variables:[102] (1) the case's circuit, or lower court, of origin; (2) issue area of the case;[103] (3) type of appellant;[104] (4) type of respondent;[105] (5)

---

[98] One programmer was able to recreate the complete works of Shakespeare at random using a few million virtual monkeys. Jesse Anderson, *A Few Million Monkeys Randomly Recreate Shakespeare*, JESSE ANDERSON BLOG (Sept. 23, 2011), http://www.jesse-anderson.com/2011/09/a-few-million-monkeys-randomly-recreate-shakespeare/.

[99] AYRES, *supra* note 5, at 10 ("Super Crunchers . . . have analyzed large datasets to discover empirical correlations between seemingly unrelated things.").

[100] *See* Ruger et al., *supra* note 4.

[101] *See* Ruger et al., *supra* note 4, at 1154–55.

[102] *Id.* at 1163. The variables the Project utilized were based on the Supreme Court database definitions, which provide coding corresponding to each variable. *Id.* at 1163 n.45. The coding is too extensive to be replicated in the footnotes, but is available on the Supreme Court Database's website. *See* HAROLD SPAETH ET AL., SUPREME COURT DATABASE CODE BOOK: 2011 RELEASE 01, at 1, 12, 14, 20, 27, 35, 44 (2011), *available at* http://scdb.wustl.edu/_brickFiles/2011_01/SCDB_2011_01_codebook.pdf.

[103] *See* SPAETH ET AL., *supra* note 102, at 35. There are too many issue areas to list (fourteen), but as an example, the ten cases discussed above have the following issue areas, as coded in the data available at SUPREME CT. DATABASE, http://scdb.wustl.edu: *Citizens United v. Fed. Elections Comm'n*, 130 S. Ct. 876 (2010) (First Amendment); *United States v. Stevens*, 130 S. Ct. 1577 (2010) (First Amendment); *Maryland v. Shatzer*, 130 S. Ct. 1213 (2010) (criminal procedure); *Johnson v. United States*, 130 S. Ct. 1265 (2010) (criminal procedure); *Padilla v. Kentucky*, 130 S. Ct. 1473 (2010) (criminal procedure); *Graham v. Florida*, 130 S. Ct. 2011 (2010) (criminal procedure); *Bloate v. United States*, 130 S. Ct. 1345 (2010) (criminal procedure); *McDonald v. City of Chicago*, 130 S. Ct. 3020 (2010) (criminal procedure); *Perdue v. Kenny A.*, 130 S. Ct. 1662 (2010) (attorneys); *Bilski v. Kappos*, 130 S. Ct. 3218 (2010) (economic activity).

[104] *See* SPAETH ET AL., *supra* note 102, at 12 (explaining the "petitioner" variable as referring to the party who petitioned the Supreme Court). Again, the types of appellants are too numerous to list (300), but for the ten cases discussed, the types of appellants as coded in the database include: *Citizens United*, 130 S. Ct. 876 (political candidate, activist, committee, party, party member, organization, or elected official); *Stevens*, 130 S. Ct. 1577 (United States); *Shatzer*, 130 S. Ct. 1213 (state); *Johnson*, 130 S. Ct. 1265 (defendant); *Padilla*, 130 S. Ct. 1473 (alien, person subject to a denaturalization proceeding, or one whose citizenship is revoked); *Graham*, 130 S. Ct. 2011 (juvenile); *Bloate*, 130 S. Ct. 1345 (defendant); *McDonald*, 130 S. Ct. 3020 (resident); *Perdue*, 130 S. Ct. 1662 (government official, or an official of an agency established under an interstate compact); *Bilski*, 130 S. Ct. 3218 (inventor, patent assigner, trademark owner or holder).

[105] *See* SPAETH ET AL., *supra* note 102, at 14. Again, the types of respondents are too numerous to list (this variable uses the same coding as types of appellants), but for the ten cases discussed, the type of respondents as coded in the database include: *Citizens United*, 130 S. Ct. 876 (Federal Election Commission); *Stevens*, 130 S. Ct. 1577 (person convicted of a crime); *Shatzer*, 130 S. Ct. 1213 (defendant); *Johnson*, 130 S. Ct. 1265 (United States); *Padilla*, 130 S. Ct. 1473 (state); *Graham*, 130 S. Ct. 2011 (state);

ideological direction of the lower court ruling (liberal or conservative, however nebulously that is defined);[106] and (6) whether the petitioner argued that a law or practice being challenged was unconstitutional.[107] The decision tree works by starting with a question, such as "Is the lower court decision liberal?," and based on whether the answers to the questions are yes or no, the tree predicts how a Justice would vote.[108]

¶57 After members of the project manually coded the value of each of these variables for all cases argued during the October 2002 Term, the model predicted the vote for each Justice based upon the decision tree.[109] The model scanned for numeric patterns rather than considering the merits of the case, a Super Cruncher algorithm that differs from the way that the experts considered cases. The trees ultimately yielded an affirm or reverse vote for each case.[110]

¶58 To test the accuracy of the decision tree model, the authors vetted and recruited a coterie of reputable Supreme Court followers. They selected these experts based on factors including writing, experience,[111] appellate practice before the Court,[112] and Supreme Court clerkships.[113] At the conclusion of the October 2002 Term, the authors compared the results from the Super Crunching decision-tree model[114] and the experts. Their model predicted 75% of the cases correctly, which was more accurate than their experts (a sample size of three) who only predicted 59.1% of the cases correctly.[115]

¶59 The authors provided a number of reasons to explain this result, such as the fact that the model predicted the economic activity cases much better than the experts.[116] The main factor, unsurprisingly, was the ability of the decision tree to predict the votes of Justices O'Connor and Kennedy.[117] The Project's authors noted that the "model seems to have captured patterns in [Justice O'Connor's] decisional behavior that the experts did not recognize."[118] Generally, lawyers can use "their experience along with traditional methods of legal analysis such as logic, analogy, and statutory interpretation to predict

Bloate, 130 S. Ct. 1345 (United States); *McDonald*, 130 S. Ct. 3020 (city, town, township, village, or borough government or governmental unit); *Perdue*, 130 S. Ct. 1662 (child, children, including adopted or illegitimate); *Bilski*, 130 S. Ct. 3218 (Department or Secretary of Commerce).

[106] *See* SPAETH ET AL., *supra* note 102, at 27.

[107] *Id.* at 44.

[108] *See, e.g.*, Ruger et al., *supra* note 4, at 1166 fig.1.

[109] *Id.* at 1163–67.

[110] *See id.*

[111] *Id.* at 1168.

[112] *Id.* at 1178 ("The practicing attorneys who participated in [the Supreme Court Forecasting Project] are appellate lawyers who appear regularly before the Supreme Court. Prediction of Supreme Court outcomes, in order to advise clients and develop litigation strategies, is an important element of their professional role."). Unfortunately, the authors did not list the identity of their "experts." The effect that lack of anonymity may have on an expert's willingness to publicly declare her predictions—and thereby open herself up to criticism if the prediction turned out to be incorrect—is unclear.

[113] *Id.* at 1168.

[114] AYRES, *supra* note 5, at 10 (noting that Super Crunching "is statistical analysis that impacts real-world decisions").

[115] *See, e.g.*, Ruger et al., *supra* note 4, at 1171.

[116] *Id.* at 1175. According to the Supreme Court Database codebook, "Economic activity is largely commercial and business related; it includes tort actions and employee actions vis-a-vis employers." SPAETH ET AL., *supra* note 102, at 34.

[117] Ruger et al., *supra* note 4, at 1172–75.

[118] *Id.* at 1173.

case outcomes for their clients."[119] However, statistical models, such as the decision tree used in the Forecasting project "often turn out to be better crystal balls than traditional experts."[120]

E. Comparing the Power Predictors and the Forecasting Project

¶60 This section compares the accuracy of the FantasySCOTUS power predictors and crowd with the Supreme Court Forecasting Project's experts and the decision tree. The power predictors predicted 64.7% of the cases correctly, surpassing the Forecasting Project's Experts (59.1%), though the difference was not statistically significant. During the October 2002 Term, the Project's model predicted 75% of the cases correctly, which was more accurate than all but the most accurate power predictors (those who had an average accuracy rate of 75.7%).

1. Of Experts and Crowds

¶61 Comparisons between the Forecasting Project's experts and the FantasySCOTUS power predictors are imprecise for several reasons. First, the FantasySCOTUS data set is derived exclusively from a crowdsourced prediction market. We did not develop any predictive model nor did we vet any experts. Unlike the "experts" selected in the Forecasting Project—mostly appellate litigators, former Supreme Court clerks, and professors—the FantasySCOTUS power predictors unknowingly selected themselves by predicting more than seventy-five percent of the cases.

¶62 When comparing the power predictors with the Forecasting Project's experts, we are not comparing two similar groups. The former is effectively a crowd, while the latter is a group of specialized experts with largely similar experience and education. Though the subset of only three members reduces the sample size, the composition of the power predictors meets the minimum size required for the statistical measures we used, and can statistically be considered a crowd.[121] Empirically, this selection approximates a normal distribution, and is still valid.

¶63 Another manner in which the power predictors differed concerned the scope of cases predicted. In the Forecasting Project, the experts were subject-matter experts—that is, they made predictions in case areas they were familiar with, such as corporate law, criminal law, constitutional law, and so on. Only three of the eighty-three experts in the Forecasting Project made predictions for most of the cases. FantasySCOTUS had a stable thirty-member cadre of power predictors that predicted a majority of the cases, from a noteworthy Second Amendment case to a less popular original jurisdiction water rights case. The power predictors' wide breadth of knowledge and experience—from Supreme Court advocate to actuary—drew from a diverse crowd with a combined wisdom that yielded a respectable accuracy rate.

---

[119] Edward K. Cheng, *Will Quants Rule the (Legal) World?*, 107 MICH. L. REV. 967, 975 (2009).
[120] *Id.*
[121] When using statistics, most measures assume a normal distribution, which is technically very rare. When dealing with groups, however, the central limit theorem states that as the sample size increases, the sample more closely approximates a normal distribution. *See* EVERITT, *supra* note 81, at 64.

2. Analysis

¶64 It was impossible to compare FantasySCOTUS' data and the Forecasting Project's data directly. The Forecasting Project looked at the October 2002 Term and FantasySCOTUS considered the October 2009 Term. There were different cases, different arguments, and, perhaps most significantly, different Justices. Indeed, the ability to utilize a significant amount of data concerning that Court's prior decisions was part of the rationale underlying the Project's model, which had an unprecedented consistent membership for nearly a decade.[122]

¶65 FantasySCOTUS data are based on the Court's October 2009 Term, which was a brand-new natural court[123] with the addition of Justice Sotomayor and the departure of Justice Souter.[124] Also, the rest of the Court's makeup had changed in the recent past, with the passing of Chief Justice Rehnquist and Justice O'Connor's retirement.[125] Pundits are still trying to figure out the Roberts Court.[126] Any benefits that either the Forecasting Project's decision tree or the experts could gain from experience about the Court likely did not exist for the participants in FantasySCOTUS. In this sense, it was likely harder to predict the October 2009 Term than the October 2002 Term.

¶66 Putting aside the temporal disparities, in Table 3 we calculated the overall accuracy ratio as a percentage—which is not dependent on specific terms, cases, or Justices—for the FantasySCOTUS power predictors and crowd, as well as the Forecasting Project's experts and decision tree.[127]

TABLE 3. ACCURACY RATIOS

Group	Correct	Incorrect[128]	Total
FantasySCOTUS Crowd	30 (44.0%)	38 (56.0%)	68 (100%)
FantasySCOTUS Power Predictors	44 (64.7%)	24 (35.3%)	68 (100%)
Forecasting Project Experts' Aggregate Votes	101 (59.1%)	70 (40.9%)	171 (100%)
Forecasting Project Decision Tree	51 (75%)	17 (25%)	68 (100%)

¶67 The FantasySCOTUS crowd performed the worst, with an accuracy of 44%. In comparison with the Forecasting Project, the results from FantasySCOTUS power

122 Ruger et al., *supra* note 4, at 1160–61.
123 "A natural court is a period during which no personnel change occurs." SPAETH ET AL., *supra* note 102, at 30.
124 *Members of the Supreme Court of the United States*, SUP. CT. U.S., http://www.supremecourt.gov/about/members.aspx (last visited Dec. 28, 2011).
125 *Id.*
126 *See e.g.*, Adam Liptak, *The Most Conservative in Decades*, N.Y. TIMES, July 25, 2010, at A1.
127 The number of cases over which we measured the wins and losses (sixty-eight) is only equal to the number used in the Forecasting Project by coincidence. For example, we removed a number of split outcomes, such as *Free Enterprise Fund v. Public Co. Accounting Oversight Board*, 130 S. Ct. 3138 (2010), which was affirmed in part and reversed in part—an outcome that FantasySCOTUS could not easily predict. We also excluded cases which were carried over to the next term, such as *Harrington v. Richter*, 130 S. Ct. 1506 (2010) (mem.).
128 For measurement purposes, cases where the same number of users predicted the case would be affirmed and reversed were treated as incorrect to avoid inflating the results. A tie is not a correct prediction.

predictors present a success story, in part. The power predictors, compared to the experts used in the Forecasting Project, predicted a higher percentage of cases correctly—64.7% to 59.1%. This 5.6% difference is not significant enough to draw any broad conclusions about the comparative expertise of the power predictors compared to the Project's experts.[129] If the two groups were to make predictions for the October 2012 Term, for example, we could not rule out the possibility that the Forecasting Project's experts would not outperform the FantasySCOTUS power predictors. These results suggest that further testing could bring the power predictors' results closer to the accuracy rate of the decision tree. The Forecasting Project's decision tree performed better than the FantasySCOTUS power predictors—75% to 64.7%. Comparatively, the FantasySCOTUS power predictors rank between the Forecasting Project's experts and the decision tree Super Cruncher algorithm.

¶68 This comparison demonstrates that in this instance the wisdom of the crowds surpassed specialized experts, yet the Super Cruncher decision tree surpassed the crowd. As Professor Ayres noted, Super Crunchers have the power of "invading and displacing traditional experts," such as the Supreme Court experts the Forecasting Project selected, and, as this analysis shows, even the wisdom of the crowds.[130]

¶69 Only three experts in the Forecasting Project accurately predicted a majority of the cases (more than 50%, thirty-five out of sixty-eight cases). In order to obtain a more accurate analysis and compare similar actors, Table 4 calculates the accuracy of predictions made by the top three Forecasting Project experts, the top three power predictors, as well as the Forecasting Project's decision tree.

TABLE 4. TOP THREE EXPERTS VERSUS TOP THREE POWER PREDICTORS

Group	Correct	Incorrect	Total
The Forecasting Project's Top Three Experts' Aggregate Votes	101 (59.1%)	70 (40.9%)	171 (100%)
The Forecasting Project's Decision Tree	51 (75.0%)	17 (25.0%)	68 (100%)
FantasySCOTUS' Top Three Power Predictors' Aggregate Votes	153 (75.7%)	49 (24.3%)	202 (100%)

¶70 The FantasySCOTUS top triumvirate, who averaged a 75.7% accuracy rate, surpassed the top three Forecasting Project experts, who averaged a 59.1% accuracy rate. These results suggest that the experts who predicted the most cases for the Forecasting Project did not have the predictive prowess the authors were seeking. It is unclear if credentials and pedigree—such as scholarship, Supreme Court practice, and Supreme Court clerkships, the metrics the Forecasting Project selected—sufficiently signal a prognosticator's jurisprudential prescience. From these two data points—the Forecasting Project and FantasySCOTUS—it appears that credentials do not correlate with an ability to predict cases. The FantasySCOTUS top three power predictors not only outperformed the Forecasting Project's top three experts but they also slightly outperformed the

[129] At a 90% confidence interval, the margin of error for the power predictors' prediction is 63.7% ± 9.53%. At the low end, the power predictors' accuracy rate is only 54.17%, lower than the Forecasting Project's experts' rate. However, at the high end, the power predictors' accuracy is 73.23%, only 1.77% away from the decision tree's accuracy.

[130] AYRES, *supra* note 5, at 11.

decision-tree algorithm—75.7% to 75%. Justin Donoho, the Chief Justice of FantasySCOTUS, achieved an accuracy rate of 80%, while the second and third place users scored 75% and 72% respectively.

¶71 The current iteration of the decision tree, however, suffers from an obvious potential defect: "[I]t cannot handle newly appointed Justices."[131] The Forecasting Project took advantage of a consistent Court, with no new appointments in nearly a decade. During the October 2002 Term the Court had been made up of the same composition of Justices for almost a decade, since Justice Breyer had joined the Court in 1994.[132] The Forecasting Project made note of the natural court's[133] stability as a "unique opportunity for research."[134] That cohort of Justices had developed a stable relationship and voting pattern. In the last seven years, there have been four new appointments, including, most significantly, a new Chief Justice and the replacement of Justice O'Connor's swing vote with Justice Alito's more predictable vote.

¶72 While the decision tree was capable of generating an impressive accuracy rate, it might not be able to serve as a viable model for predicting Supreme Court cases year after year with a changing Court. "Even if a model could be constructed that perfectly fit past Supreme Court outcomes, we could not be certain that the model's variables and their relationships would remain useful over time."[135] In contrast, a crowdsourced model is flexible, resilient, and self-evolving. The members of the prediction market naturally take note of the changes in perceptions of the Justices, and adapt accordingly. If Justice Ginsburg were to retire next term, for example, the members may have some uncertainty as to how her successor will vote, but the market would still continue. Further, FantasySCOTUS does not rely on the manual categorization of cases—a subjective process that could insert biases and undue influences into any research. It will be possible in the future to use sophisticated algorithms to Super Crunch data from cases based on precedents, judicial philosophy, and rules of law, rather than on the voting history of a specific set of Justices. This methodology will allow for the prediction of cases, with any composition of Justices or judges, in any court. This evolution will enable the development of sophisticated judicial prediction engines.[136]

V. LIMITATIONS AND VALUE OF FANTASYSCOTUS 1.0

¶73 The value of the first season of FantasySCOTUS for the October 2009 Term, or FantasySCOTUS 1.0 as we call it, is quite modest, and should be kept in perspective. The FantasySCOTUS power predictors—those who made predictions for more than

[131] Samaha, *supra* note 31, at 834 (citing Ruger et al., *supra* note 4, at 1169–70, 1170 n.67).
[132] Ruger et al., *supra* note 4, at 1154; *see also Members of the Supreme Court of the United States*, *supra* note 124. This makeup of the Court lasted until the death of Chief Justice Rehnquist and the appointment of Chief Justice Roberts on September 29, 2005. This cadre of the Rehnquist Court, which lasted eleven years, is tied with the 1812–1823 Marshall Court for the longest group of Justices to serve together.
[133] Ruger, et al., *supra* note 4, at 1160 n.38 ("We adopt the commonly accepted definition of 'natural court' as referring to a period of time where the same nine Justices sit together on the Supreme Court without any composition change.") (citing JOAN BISKUPIC & ELDER WITT, THE SUPREME COURT AT WORK 315 n.a (2d ed. 1997)).
[134] *Id.* at 1160.
[135] Samaha, *supra* note 31, at 835.
[136] *See infra* Part VII-C (discussing the evolution of FantasySCOTUS).

seventy-five percent of the cases—were accurate in 64.7% of their predictions. The top three power predictors in FantasySCOTUS scored accuracy rates of 80%, 75%, and 72%, respectively (an average of 75.7%). As novel as these results are for the first season, FantasySCOTUS' current predictive capabilities are respectable, but not reliable—it was wrong, in the best case scenario, between 25% and 35% of the time.

¶74 Further, the Supreme Court typically reverses about 70% of the cases it decides each term.[137] During the October 2009 Term, for example, the Court reversed 72% of the cases decided on the merits.[138] In theory, predicting that the Supreme Court would reverse for every case would have yielded a 72% accuracy rate, and a top-three finish (we verified that no member of FantasySCOTUS made predictions in this manner).

¶75 FantasySCOTUS 1.0 should be understood for what it does and does not do. The authors of the Forecasting Project recognized that "[w]hat is notable, in light of all the attention focused on the Court, is that few have tried to systematically predict its decisions prospectively."[139] In its current iteration, FantasySCOTUS provides real-time *ex ante* predictions for individual cases. No other product performs this task for every case argued during the Term. Comparing these results to *ex post* aggregate analysis, such as the overall reversal rate of 72%, is imprecise. Simply concluding *ex post* that the Court reversed approximately 72% of all cases argued during a term provides no information about individual cases.

¶76 In contrast, FantasySCOTUS generated real-time predictions for every pending case—not just an aggregate overall prediction of what could have been after the Term. Further, the 72% reversal rate provides no information about which 72% of the docket will be reversed. The reversals do not necessarily occur during the first or last cases decided and are distributed throughout the Term, with the reversal granted based on the merits of the case, not the remaining number of cases and outcomes.

¶77 To put it another way, armed solely with the 72% aggregate reversal rate, a predictor would have no way *ex ante* of knowing how an individual case will turn out. To say that any individual case has a 72% likelihood of reversal is a statistical fallacy. One would have to know the specifics of the case to make that type of estimate. Comparing *ex post* aggregate trends and *ex ante* predictions fails to account for the independence of each case.

¶78 Additionally, the power predictors' accuracy rates of 65% to 75% consist of data points for each case, with an attendant confidence level of at least 90%, or in some cases 95% or 99%. For many of the 25% to 35% of cases that FantasySCOTUS failed to accurately predict, we knew *ex ante* that we lacked sufficient data to make an accurate prediction. For the most part, we were not surprised when the predictions were correct.

---

[137] *See, e.g.*, *Stat Pack OT2010*, *supra* note 6, at 4 (indicating the Supreme Court reversed 72% of its cases during the October Term 2010); *Stat Pack OT2009*, *supra* note 6, at 10 (indicating the Supreme Court reversed 71% of its cases during the October Term 2009); *SCOTUSBlog Stat Pack Final Data 6.29.09*, SUP. CT. U.S. BLOG 10 (June 29, 2009), http://www.scotusblog.com/wp-content/uploads/2009/06/full-stat-pack.pdf (indicating the Supreme Court reversed 75.9% of its cases during the October Term 2008).

[138] *Final Stats OT09–7.7.10*, *supra* note 6, at 2.

[139] Ruger et al., *supra* note 4, at 1154. Also notable is how little attention is paid to attempting to accurately catalogue the past work of the Court, which can be crucial for determining how the Court or individual Justices may resolve a case in the future. *See, e.g.*, Ross E. Davies, Craig D. Rust & Adam Aft, *Justices at Work, or Not: New Supreme Court Statistics and Old Impediments to Making Them Accurate*, 14 GREEN BAG 2D 217 (2011).

Likewise, we were not surprised when the predictions were incorrect, based upon the standard statistical measures of reliability we were able to generate based on the predictions before the Court decided the case.

¶79 Consider two cases decided during the October 2009 Term. For *Levin v. Commerce Energy, Inc.*,[140] the data were not significant, and would not yield an accurate prediction. For *Wood v. Allen*,[141] the data were significant, and we were virtually certain that the prediction would likely be accurate. For *Levin v. Commerce Energy*, only fifty-five percent of members predicted that the Supreme Court would affirm the lower court. At a 90% confidence level, the confidence interval was ± 11.57.[142] Thus, the actual likelihood of an affirmance could be as high as 66.57%, or as low as 43.43%. If the likelihood of an affirmance reaches below 50%, we can no longer be confident that the prediction will be accurate. At the time, in a *Predictions of the 10th Justice* column, we noted that the data were "not strong enough for the [prediction] to be significant."[143] The Court ultimately reversed the lower court 9–0, a minority (forty-five percent) correctly predicting a reversal.

¶80 In contrast, *Wood v. Allen* provides an instance where we knew *ex ante* that our predictions would almost certainly be accurate.[144] In the case, 80% of members predicted that the Court would affirm the Eleventh Circuit. At a 99% confidence level, the confidence interval was ± 11.65.[145] The actual likelihood of an affirmance could have been as low as 68.35%, or as high 91.65%. In either scenario, the confidence at a 99% confidence level that the Court would affirm was greater than 50%. Recognizing this certainty, at the time, we noted that FantasySCOTUS members would be "extremely accurate at predicting the general outcome."[146] We were not surprised when the Justices voted to reverse. With the appropriate confidence interval, we can signal in advance the statistical measures indicating whether the prediction stays above 50% for affirm or reverse, and, if so, the confidence interval at which that prediction stays above 50%. From this, we can determine how reliable, or unreliable, a prediction is.

¶81 While a thirty-five percent failure rate is still largely unhelpful—it is doubtful anyone could meaningfully rely on FantasySCOTUS at its present accuracy rate—a larger subscriber base could increase the accuracy. FantasySCOTUS 1.0 had 5,000 members. FantasySCOTUS 2.0—the season that began with the October 2010 Term—has approximately 10,000 members. With developing partnerships with Westlaw and enhanced marketing strategies, we hope to double that number next season. As FantasySCOTUS grows, and more members sign up, with a wider range of views and opinions, our crowd grows, and we can obtain more data points. With more data points, the confidence interval shrinks. Even at higher confidence levels (90%, 95%, and 99%), we expect to see more reliable predictions above 50% to either affirm or reverse.

---

[140] 130 S. Ct. 2323 (2010).
[141] 130 S. Ct. 841 (2010).
[142] Josh Blackman, *FantasySCOTUS.net—Revisiting* American Needle, Graham v. Florida, Comstock, *and* Berghuis, JOSH BLACKMAN'S BLOG (June 2, 2010, 9:35 PM), http://joshblackman.com/blog/?p=4577.
[143] *Id.*
[144] Josh Blackman, *FantasySCOTUS.net Predictions of the 10th Justice: Testing the Wisdom of the Crowds*, JOSH BLACKMAN'S BLOG (Feb. 4, 2010, 10:46 PM), http://joshblackman.com/blog/?p=3951.
[145] *Id.*
[146] *Id.*

¶82 FantasySCOTUS 1.0 generated a data set that allowed us to develop an analytical framework to devise a prediction market for the Supreme Court. FantasySCOTUS also provides new insights into predicting Justice behavior. As we continue to gather data, we can see what this information teaches us about the models of judicial decision making, and whether applying different models (such as attitudinal models[147]) yields different types of predictions. In learning about how people predict the Justices will interact, we may learn something about how they actually interact and thus something about the institution of the Court itself.[148] As this technology develops in the future, possibilities of an automated approach to understanding judicial behavior are vast.

VI. Impact of the Media

¶83 FantasySCOTUS, and Supreme Court speculation in general, may pose somewhat of a chicken and egg problem. Are predictions of members organically developed based on the existing precedents and how the Justices interact during oral arguments? Or, do media accounts that describe these precedents and interactions artificially generate predictions in the minds of members? In other words, does FantasySCOTUS "react more than [it] predicts"?[149] This section explores the relationship between the nature of Supreme Court predictions and the impact media coverage plays in these predictions. More specifically, we focus on the benefits of a prediction market, even in light of its potential reactionary tendencies.

¶84 Excluding two outlier cases, we found a strong correlation between the amount of media attention and the accuracy of predictions for both the power predictors and the crowd. The power predictors' edge is dulled for cases that receive significant media coverage. For unpopular cases that receive less media attention, and thereby less information for prospective predictors, the crowd tends to generate less accurate predictions. In contrast, the power predictors, who perform their own due diligence and research irrespective of media coverage, can thrive even on the most obscure cases on the docket.

A. Reactionary Prediction Markets

¶85 Professor Orin Kerr posed an interesting question about TradeSports, a prediction market that aimed to predict who President Bush would nominate to replace Justice O'Connor. The morning that President Bush announced that Judge John Roberts was his nominee, TradeSports erroneously predicted that Judge Edith Clement—the popular nominee in most media accounts—would be the nominee.[150] Presaging this faulty pick based on media consensus, Kerr wrote:

> The choice of O'Connor's replacement belongs to one man, George W. Bush. A few inside advisors are privy to his thinking, but I think it's fair to assume that neither Bush nor any of his inside advisors are placing any bets on sites like

[147] *See* Maxwell L. Stearns & Todd J. Zywicki, Public Choice Concepts and Applications in Law 406–97 (2009).
[148] *See id.*
[149] Note, *supra* note 15, at 1223.
[150] *See* Abramowicz, *supra* note 91, at 38–39.

> TradeSports.com. This means that the people who are placing bets presumably are outsiders who are getting their predictions from newspaper articles, blogs, horoscopes, etc., and then placing bets. As a result, a site like TradeSports would seem to just mirror the collective common wisdom of newspapers and blogs on a question like this. Am I missing something?[151]

Commenting on Kerr's post, Michael Abramowicz noted that prediction markets "do not seem to tell participants much more than they could figure out themselves by considering the underlying materials."[152] Do prediction markets merely repeat information in the media?

¶86 FantasySCOTUS provides a unique opportunity to test Professor Kerr's idea. In order to consider the impact of media coverage on all of the participants—the crowd and power predictors—we gathered data on the media attention of each case and compared it against the accuracy of predictions. This data helped to answer two questions: first, did the media attention the cases received correlate with the accuracy of the predictions; and second, can the cases that had a statistically significant gap between the crowd's predictions and power predictors' predictions be explained in part by the quantum of media attention?

¶87 Even if prediction markets are primarily reactive, they still serve a very important role—aggregation. How can one quantify the "collective common wisdom of newspapers and blogs"?[153] Markets, such as FantasySCOTUS, serve as a clearinghouse of sorts and provide an easy way to assemble the totality of knowledge in the media and elsewhere, even if the predictions merely reflect that consensus. The prediction markets "at least opened up the possibility of accomplishing the task of evidence aggregation,"[154] which is a very important task. On average, the prediction market will "be more accurate than the prediction that the observer independently"—even Professor Kerr—"could derive, because the market will represent an aggregation of the views of a large number of observers."[155] Specifically, this aggregation may "in part cancel out random errors that individuals make in predictions by overweighing or underweighing particular pieces of evidence."[156]

¶88 Like unfounded guesses as to who will be the next Supreme Court Justice, where the knowledge is likely restricted to a few people in the Executive Branch, the outcome of Supreme Court cases are known only by the Justices and their clerks. There is no special inside information, known to reporters and supposed experts. Rather, the wise crowd, who are able to read the tealeaves and pick out important questions asked by the Justices, can determine how the Court will decide.

¶89 Some research claims that "prediction markets will work better when they concern events that are widely discussed, since trading on such events will have higher entertainment value and there will be more information on whose interpretation traders can disagree."[157] Even if the information surrounding a case is "ambiguous," perhaps

---

[151] Kerr, *supra* note 7.
[152] ABRAMOWICZ, *supra* note 91, at 38.
[153] Kerr, *supra* note 7.
[154] ABRAMOWICZ, *supra* note 91, at 38.
[155] *Id.*
[156] *Id.*
[157] Wolfers & Zitzewitz, *supra* note 27, at 121.

resulting from contentious oral arguments or a longer-than-usual delay in issuing the opinion, this data may result in better predictions because those most skilled in reading between the lines and figuring out the Justice's proclivities can put forth the best guesses.[158]

B. Methodology

¶90 In order to assess whether media coverage and prediction accuracy are correlated, the extent of press coverage about a case is relevant. However, other than limited survey data from some power predictors,[159] we had no basis to determine which media sources FantasySCOTUS members relied on to learn about the cases. To solve this problem, we devised an approach to determine media coverage of Supreme Court cases.[160] To reflect the transformation of how Supreme Court cases were covered, we considered two sources of media—old school and new school. First, we looked at coverage in what could broadly be referred to as mainstream media. For this search, two comprehensive databases were utilized: the All News Plus database on Westlaw[161] and the All Legal U.S. News database on LexisNexis.[162] Utilizing sources from both Westlaw and LexisNexis increased the data set's inclusiveness and allowed for normalized results.

¶91 Years ago, to learn about a Supreme Court case, one would have to wait for Linda Greenhouse's article in the New York Times the next day or catch Nina Totenberg's spot on National Public Radio. Thanks to the legal blogosphere revolution, that is no longer the case. Many blog authors post instant analyses of oral arguments, opinions, and other developments at the Court within minutes of the breaking news.[163] To consider the impact of coverage in the blogosphere—and sort through the tangled World Wide Web—we searched the "Blogs on Demand" database in Westlaw.[164] This blog database is quite limited and excludes a number of the most popular legal blogs.[165] To focus on the

---

[158] *Id.* ("Ambiguous public information may be better in motivating trade than private information, especially if the private information is concentrated, since a cadre of highly informed traders can easily drive out the partly informed, repressing trade to the point that the market barely exists.").

[159] For a discussion how several power predictors made their decisions, and which resources they relied on, see Blackman, *supra* note 144.

[160] Objectively discerning the media attention given to a Supreme Court case is not an exact science, and there may certainly be room to improve on the method employed.

[161] WESTLAW, WINNING RESEARCH SKILLS: PROFESSIONAL LEGAL RESEARCH 185 (2008), *available at* http://lscontent.westlaw.com/images/banner/SurvivalGuide/PDF08/08WinningResearchSkills.pdf (The All News Plus Wires database (ALLNEWSPLUS) "contains newspapers, magazines, journals, newsletters, government press releases, and transcripts of television and radio shows and congressional testimony . . . plus newswires.").

[162] *Legal US News, All*, LEXISNEXIS, http://w3.nexis.com/sources/scripts/info.pl?7596&GCC (last visited Dec. 28, 2011) ("The ALLNWS file is a group file containing all of the separately searchable online legal newspaper, magazine, and newsletter files.").

[163] *See, e.g.*, Josh Blackman, *Instant Reaction:* Citizens United v. FEC *(Hillary Movie Case)*, JOSH BLACKMAN'S BLOG (Jan. 21, 2010, 10:08 AM), http://joshblackman.com/blog/?p=3771.

[164] *West E-lert Newsletter: Does Westlaw Provide Access to Any Legal or Financial Blogs?*, WESTLAW (May 2009), http://store.westlaw.com/signup/newsletters/westelert/2009-may/article8.aspx ("[T]he Blogs on Demand database (BLOGSOD) . . . contains the full text and abstracts of a variety of financial, legal, and business blogs . . . .").

[165] Among others, it excluded BALKINIZATION, http://balkin.blogspot.com/ (last updated Dec. 4, 2011); INSTAPUNDIT, http://pjmedia.com/instapundit/ (last updated Dec. 4, 2011); PRAWFSBLAWG,

sources that Court-followers read most closely, we augmented the search field to include the blogs listed in the "ABA Journal 4th Annual Blawg 100,"[166] which contains "the best legal blogs as selected by the [*ABA*] *Journal*'s editors" and includes "the best and brightest law bloggers in a variety of categories."[167] To comb through these sources, we programmed a custom Google search engine that searched these 100 sites.[168]

¶92 The primary search problem relates to the inconsistent ways authors refer to cases. Take *Citizens United v. Federal Elections Commission*, for example. Some authors call it *Citizens United v. FEC*, others simply *Citizens United*, and still others call it the "Hillary Movie Case." Simply searching for one of these phrases would exclude a number of relevant articles and posts. Further, a case with a common name, such as *United States v. Stevens*, which can be abbreviated as simply *Stevens*, generated a significant number of false negatives, especially in light of the newsworthy retirement of Justice John Paul Stevens.

¶93 To minimize inaccurate results, we went through a process of testing multiple search strings and reviewing the results to determine which terms would be most accurate and allow the greatest consistency.[169] Using just the unique party name in a case worked relatively well for *Citizens United*, but we could not replicate that success with cases such as *Stevens* or *Johnson*. In fact, when we just ran the more unique party name, six of the ten cases hit 10,000 search results on the Westlaw All News database,[170] indicating that we exceeded the maximum size of search results permitted on that database.[171]

¶94 Utilizing the proximity searches was not particularly helpful; they were either over-inclusive and maximized the search results on the databases (for example, searching for terms in the same sentence), or under-inclusive and did not return a noticeably greater amount of search results (for example, searching for terms within two words of each other). Ultimately, we ran straightforward search strings for each of the cases in all four databases (for example, "citizens united v. federal election commission"). This strategy was under-inclusive for cases such as *Citizens United*, where many commenters did not use the full case name. However, the strategy that provided the most accurate results was

---

http://prawfsblawg.blogs.com/ (last updated Dec. 4, 2011); WALL ST. J.L. BLOG, http://blogs.wsj.com/law/ (last visited Dec. 4, 2011).

[166] Molly McDonough & Sarah Randag, *Our 100 Favorite Blawgs*, A.B.A. J., Dec. 2010, at 33. In the interest of full disclosure, *Josh Blackman's Blog* was selected to this list. *Id.* at 34.

[167] *The 2010 ABA Journal Blawg 100*, A.B.A. J., http://www.abajournal.com/blawg100 (last visited Nov. 13, 2011).

[168] *JoshBlackman.com ABA Blawg 100 Search*, GOOGLE CUSTOM SEARCH, http://www.google.com/cse/home?cx=003923726555708584283:mzan1oszugg (last updated Mar. 7, 2011). *See generally* GOOGLE CUSTOM SEARCH, http://www.google.com/cse/ (last visited Dec. 28, 2011).

[169] A few examples of attempted searches include—with *Citizens United* as an example—["citizens united" /s "federal election commission"]; ["citizens united" /2 "federal election commission"]; ["citizens united"]; and ["citizens united v. federal election commission"]. The search strings tested are too numerous to list. To craft searches that yielded the most accurate search results, we utilized terms and connectors searching in Lexis and Westlaw, with options such as: & (both search terms); or (either search term or both terms); " " (search terms appearing in the same order as in the quotation marks); /n (search terms within n terms of each other (where n is a number from 1 to 255); /s (search terms in the same sentence); and /p (search terms in the same paragraph). WESTLAW, WESTLAW QUICK REFERENCE GUIDE: GETTING STARTED ON WESTLAW 5 (2009), http://store.westlaw.com/documentation/westlaw/wlawdoc/web/rswlcm06.pdf.

[170] *See supra* note 161 and accompanying text.

[171] The six cases that exceeded the maximum results were *Stevens*, *Johnson*, *Padilla*, *Graham*, *McDonald*, and *Perdue*.

to search for the more common name cases, and we chose to follow the most consistent path.

¶95 To further improve the accuracy of the searches, we utilized the date restriction features in Lexis and Westlaw. The custom Google search engine we programmed did not allow for date restrictions. Attempting to focus on the media attention that a participant in FantasySCOTUS would have, we limited the date range from the date the Supreme Court granted certiorari to the date the Supreme Court issued an opinion. We compiled all of the date ranges using the Supreme Court's official docket.[172]

C. Analysis

¶96 Table 5 displays the number of results we found in each database for the keyword search between the date certiorari was granted and the date of the opinion. In the final column, we averaged the results.

TABLE 5. NEWS DATABASE RESULTS

Search	**Date Range**	**ALLNWS**	**ALLNEWS PLUS**	**BLOG SOD**	**BLAWG 100**	**Average**
"citizens united v. federal election commission"	8/18/2008[173] –1/21/2010	26	290	5	80	100.25
"united states v. stevens"	4/20/2009–4/20/2010	19	47	3	59	32
"maryland v. shatzer"	1/26/2009–2/24/2010	7	28	2	37	18.5
"johnson v. united states"	2/23/2009–3/2/2010	13	12	0	32	14.25
"padilla v. kentucky"	2/23/2009–3/31/2010	12	38	1	69	30
"graham v. florida"	5/4/2009–5/17/2010	16	88	5	83	48
"bloate v. united states"	4/20/2009–3/8/2010	2	3	0	11	4
"mcdonald v. chicago"	9/30/2009–6/28/2010	31	212	24	94	90.25
"perdue v. kenny"	4/6/2009–4/21/2010	50	52	4	26	33
"bilski v. kappos"	6/1/2009–6/28/2010	113	116	27	92	87

¶97 According to the data, *Citizens United*, *McDonald*, and *Bilski* received the greatest average media coverage, while *Bloate*, *Johnson*, and *Shatzer* received the least. To ascertain whether a correlation exists between the media coverage and the accuracy of FantasySCOTUS predictions, it is helpful to scatter plot these data with a best-fit line.

---

[172] *Docket Search*, SUP. CT. U.S., http://www.supremecourt.gov/docket/docket.aspx (last updated Dec. 29, 2011).

[173] This is the date the case was docketed, which we used in lieu of the original date the Supreme Court granted certiorari during the October 2008 Term.

¶98 Figure 2 considers the relationship between the media coverage and the accuracy of the FantasySCOTUS power predictors. Figure 2 plot considers the relationship between the media coverage and the accuracy of the FantasySCOTUS crowd.

¶99 FIGURE 1

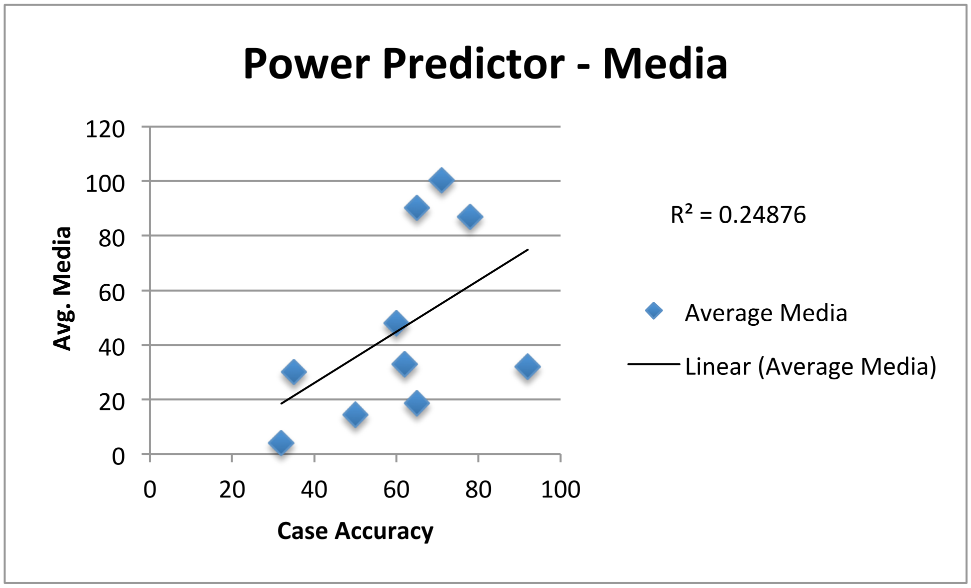
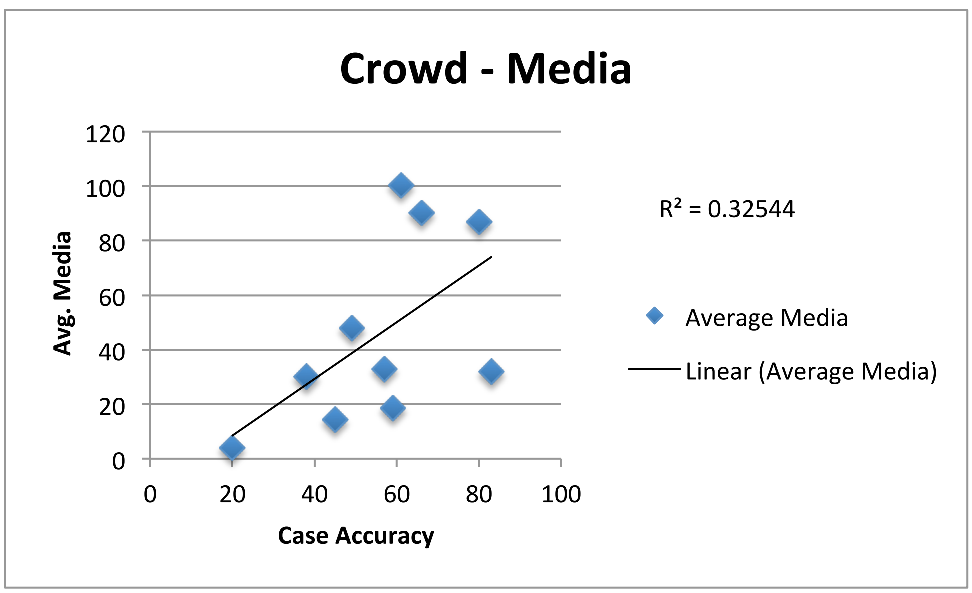


FIGURE 2



¶100 These scatter plots allow us to draw several conclusions. First, the points slope upward, suggesting that the more media attention a case received, the more accurate the predictions were. However, both the crowd and power predictor best fit lines have very low R^2 values.[174] R^2 ranges between zero and one. As the R^2 value approaches one, we can conclude that the predicted value is closer to the actual value. In other words, the plot has a higher predictive power. As the R^2 value approaches zero, the predictive power decreases, and we cannot say with confidence that the predicted value approximates the actual value. The R^2 values of 0.248 for the power predictors and 0.325 for the crowd are quite low. These values signify that the predictive power of this plot is fairly weak.

¶101 However, the crowd trended more closely with the media coverage attention than the power predictors did. Simply put, the accuracy of the crowd improved more with greater media attention relative to the power predictors. The power predictors, in contrast, were able to accurately predict cases even if the media coverage was lacking. Perhaps, the power predictors performed additional research—several members in the survey revealed that they read oral argument transcripts, the briefs, and amici[175]—to hone their results. The crowd, which predicted fewer cases, and likely invested less time into FantasySCOTUS, was perhaps more lackadaisical with their due dilligence, and merely relied on media accounts to form their votes.[176] This would seem to bolster Professor Kerr's theory that prediction markets "do not seem to tell participants much more than they could figure out themselves by considering the underlying materials."[177]

---

[174] R^2 is a statistical measurement, which represents the difference between the actual outcome and the expected outcome, in this case how far away the accuracy of a prediction was based upon the average media hits. EVERITT, *supra* note 81, at 78 (defining coefficient of determination); *see also* Mohan P. Rao & Christian D. Tregillis, *Econometric Analysis*, *in* LITIGATION SERVICES HANDBOOK: THE ROLE OF THE FINANCIAL EXPERT 6.11 (Roman L. Weil et al. eds., 4th ed. 2007) ("R^2 is a summary measure of the goodness of fit of the fitted regression line to a set of data. Formally, R^2 is defined as the ratio of explained sum of squares (variation of estimated Y values about their mean) to total sum of squares (total variation of Y values about their sample mean). R^2 ranges from 0 to 1, where 0 reflects that no variation in the dependent variable is explained by the independent variables and 1 reflects that all of the variation in the dependent variable is explained by the independent variables. Because of the heuristic simplicity of R^2, it is a widely used measure of the goodness of fit of the least squares model. . . . [T]he addition of variables to a model generally will increase its R^2. But a model with a large number of variables and a higher R^2 does not necessarily provide additional understanding of the relation between the key variables of interest and the dependent variable. . . . Further, a model with a large number of variables is harder to interpret.").

[175] Blackman, *supra* note 73.

[176] In Table 6, the difference between the correlations is much smaller, signifying that this observation is potentially attributable to the outliers. Future testing may resolve this quandary.

[177] ABRAMOWICZ, *supra* note 91, at 38.

TABLE 6. AVERAGE MEDIA AND DIFFERENCE FROM POWER PREDICTOR AND CROWD ACCURACY

Case	Average Media	Power Predictor		Crowd	
		Accuracy	Difference	Accuracy	Difference
Citizens United	100.25	71	29.25	61	39.25
Stevens	32	92	60	83	51
Shatzer	18.5	65	46.5	59	40.5
Johnson	14.25	50	35.75	45	30.75
Padilla	30	35	5	38	8
McDonald	90.25	65	25.25	66	24.25
Graham	48	60	12	49	1
Bloate	4	32	28	20	16
Perdue	33	62	29	57	24
Bilski	87	78	9	80	7
Average	-	-	15.275	-	10.075
Average, omitting outliers	-	-	5.78125	-	1.15625

¶102 The two cases with the greatest difference between media attention and prediction accuracy were *Stevens* (a difference of 60 for the power predictors and 51 for the crowd) and *Shatzer* (a difference of 46.5 for the power predictors and 40.5 for the crowd), as shown in Table 6. These cases were effectively statistical outliers. With only ten cases, the impact of these outliers significantly impacted the value of R^2, and the predictive power of the data.

¶103 *Stevens* was particularly problematic because Justice John Paul Stevens, who coincidentally shares a surname with the respondent in *Stevens*, was still sitting on the Court during the October 2009 Term. Therefore, any attempt to utilize a proximity search, such as [“united states” /s Stevens], would return a number of results talking about the United States and Justice Stevens, but not the desired search, *United States v. Stevens*. Compounding this problem was Justice Stevens’s retirement, which greatly increased the media attention he received.

¶104 *Shatzer* is a more unique party name, so it is likely that searching the full case name was somewhat under-inclusive, leading to lower media attention than actually existed. Given that we are only reviewing the data on ten cases, outliers have a much greater impact on any potential trends, and excluding them provides a more accurate picture of any potential correlation between media attention and the accuracy of any predictions.

¶105 We conducted a separate experiment omitting these outliers. Excluding *Stevens* and *Shatzer*, the average difference for the other eight cases dropped drastically: from 15.275 to 5.78 for the power predictors, and from 10.075 to 1.16 for the crowd. Without these cases, we generated two new scatter plots using the same methodologies. FIGURE 3 considers the relationship between the media coverage and the accuracy of FantasySCOTUS power predictors, excluding *Shatzer* and *Stevens*. Figure 4 considers the relationship between the media coverage and the accuracy of FantasySCOTUS crowd, excluding *Shatzer* and *Stevens*.

¶106 FIGURE 3

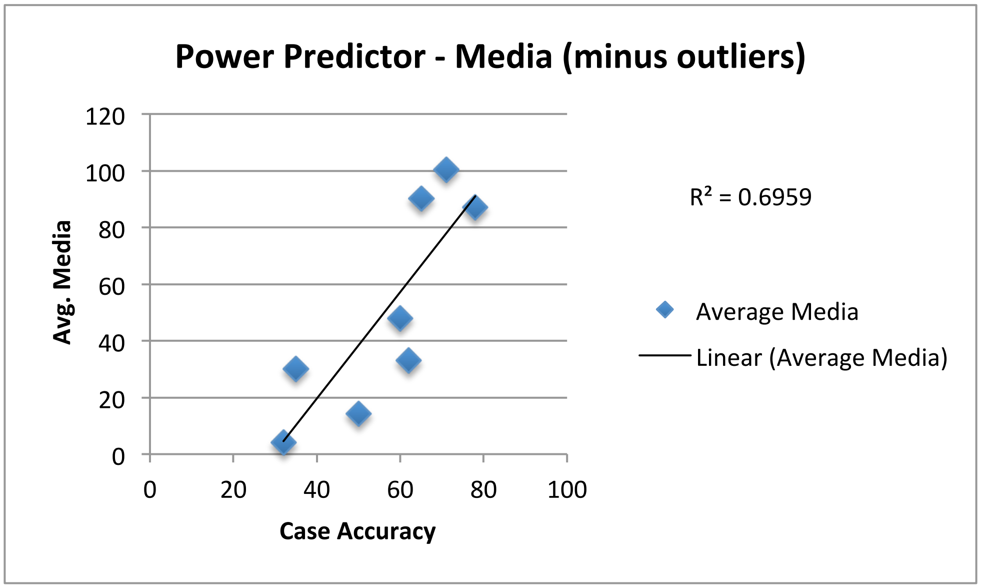
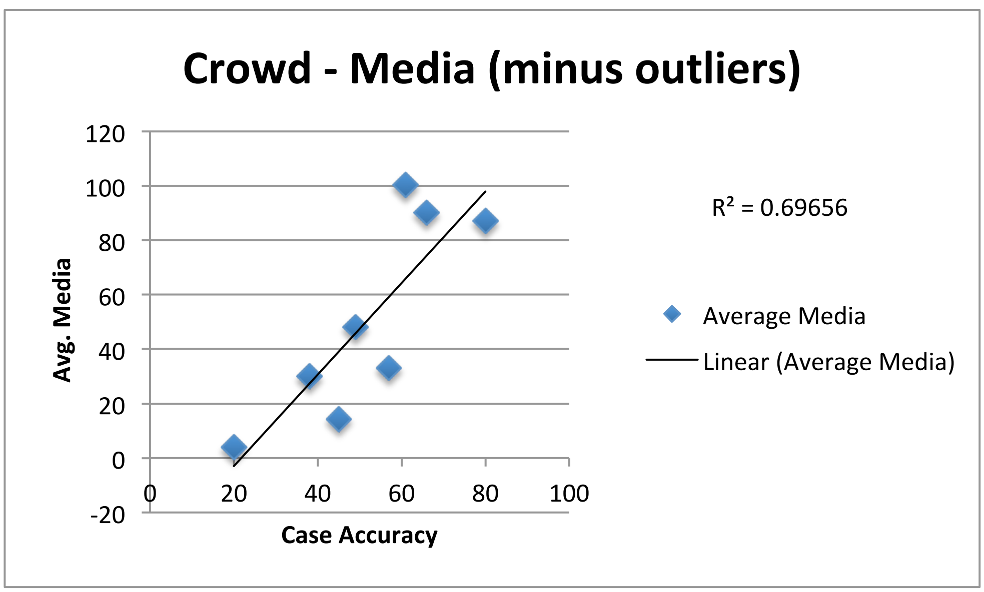


FIGURE 4



¶107 Excluding the outliers, there is a much stronger correlation between the amount of media attention and the accuracy of predictions: there was almost a three-fold increase in the correlation for power predictors between prediction accuracy and media coverage—from 0.25 to 0.69. Further, the correlation is much more similar when comparing the power predictors with the crowd—a difference of only 0.00066. An R^2 value of approximately 0.7, with only eight data points—a relatively small sample size—suggests a relatively strong correlations. The power predictors predicted both of the outlier cases more accurately. Removing those cases narrows the gap between the power predictors and the crowd in terms of correlation.

TABLE 7. AVERAGE MEDIA FOR CASES WITH SIGNIFICANT DIFFERENCES BETWEEN POWER PREDICTORS AND CROWD ACCURACY

Case	**Average Media**
Shatzer	18.5
Johnson	14.25
Graham	48
Bilski	87

¶108 There are relatively few data points using only these ten cases, so analyzing the correlation between media coverage and accuracy is mostly for observational purposes and does not have statistical significance. For comparison, given the sharp drop off in media attention from the first three cases to the last ten, this section considers the top three attention getters compared to the remainder of the cases. Three of the four cases in which the difference between the power predictors' predictions and the crowd's predictions were statistically significant were also the three cases which received the least media attention: *Johnson*, *Shatzer*, and *Graham*, ranked ninth, eighth, and fourth, respectively. The fourth case with a significant difference between power predictor and crowd predictions, *Bilski*, received significant media attention, ranking third out of the ten cases.[178] For *Bilski*, the fact that the power predictors' predictions were still significantly different than the crowd's predictions appears to be an outlier that may be explained by the highly technical nature of the patent case.

¶109 Less media attention, and thereby less information for prospective predictors in the crowd, helps to explain the crowd's weaker performance for the less-noteworthy cases. In contrast, the power predictors, who likely perform their own due diligence irrespective of media coverage, can thrive even on the most obscure and neglected cases on the docket. With only four cases with a statistically significant difference, there are too few data points to consider any correlation, but this consideration may yield interesting results in future experiments.

¶110 In short, Professor Kerr's thesis accurately observes how prediction markets generate results, but it overlooks an important aspect of these markets. Prediction markets serve a valuable function of aggregating and sorting knowledge and opinions in a unified clearinghouse. This sorting, accomplished independently by the crowd and aggregated by FantasySCOTUS algorithms, is far easier and more accurate than manually combing through and reading the unbounded amount of information printed about every

[178] Even though *Graham* had the fourth highest average of media hits and *Bilski* had the third, the jump from third to fourth, a difference of thirty-nine, is the largest jump in the average media hits. The next largest difference is only a fifteen point jump from *Perdue* to *Graham*, and the average difference from case to case is only ten points. Thus, *Graham* is more appropriately grouped with *Johnson* and *Shatzer* in this instance than with *Bilski*.

case in the mainstream media and on blogs. FantasySCOTUS tapped a "diversity of information [that] exist[ed] in a way that provides a basis for" predictions.[179]

VII. IMPLICATIONS OF FANTASYSCOTUS

¶111 In this section, we discuss some of the jurisprudential implications of creating a market that accurately predicts how the Court and, more specifically, how nine Justices will vote. This market may have a broader impact on public perceptions of the Court as an institution and its role in furtherance of the rule of law. This Part takes a brief excursion to the future and explores how a sophisticated prediction market can contribute to the evolution of the legal profession.

A. FantasySCOTUS and the Supreme Court as an Institution

¶112 FantasySCOTUS provides insights into how the Supreme Court is perceived. While the Supreme Court enjoys a favorability rating of roughly 60%, higher than the other two branches of the federal government, a recent Pew Research Center report suggests that 46% of those surveyed think the Court is too partisan—23% thinks it is liberal, 23% thinks it is conservative.[180] Only 39% believes the Court is "middle of the road."[181] Some polling exists as to how the public perceives certain noteworthy cases—for example, 68% of those surveyed disagreed with *Citizens United*[182]—but no data exist as to how people perceive the individual Justices, and their ideologies. Considering that 72% of those surveyed could not identify the name of the Chief Justice of the United States,[183] such polling of the public at large would probably be impossible, if not futile. In this sense, FantasySCOTUS serves as one of the most comprehensive, albeit unscientific, polling mechanisms to capture perceptions of the Supreme Court as an institution.

¶113 While FantasySCOTUS 1.0 did not request that members identify their ideology—we requested this information in version 2.0, and hope to elaborate on this dynamic in future work—anecdotal evidence suggests that certain members consistently voted in a manner that reflected a particular jurisprudential ideology. This bias usually manifests in predictions for Justice Kennedy's often decisive vote in 5–4 decisions. Members who voted for outcomes that could be deemed liberal would align Justice Kennedy's vote with those of Justices Stevens, Ginsburg, Breyer, and Sotomayor. Many of these same members made outlier predictions in a related FantasySCOTUS game to predict Justice Stevens's replacement, selecting long shots like Cass Sunstein or Pam Karlan. These votes were likely based on their personal predilections rather than realistic expectations. In contrast, members who voted for outcomes that could be deemed conservative would

---

[179] Wolfers & Zitzewitz, *supra* note 27, at 120.
[180] *The Invisible Court*, PEW RES. CENTER (Aug. 3, 2010), http://pewresearch.org/pubs/1688/supreme-court-lack-of-public-knowledge-favorability.
[181] *Id.*
[182] News Release, Pew Research Ctr., Obama's Ratings Are Flat, Wall Street's Are Abysmal: Midterm Election Challenges for Both Parties 31 (Feb. 12, 2010), *available at* http://www.people-press.org/files/legacy-pdf/589.pdf.
[183] News Release, Pew Research Ctr., Political Knowledge Update 2 (July 15, 2010), *available at* http://people-press.org/files/legacy-pdf/635.pdf.

align Justice Kennedy's vote with those of Chief Justice Roberts, and Justices Scalia, Thomas, and Alito.

¶114 By enabling people to vote their preferences, and thereby express their views of how ideological cases will be decided, FantasySCOTUS can gather how people view the Justices. This collective wisdom of the crowds captures the public coarsening among lawyers and law students—the vast majority of players on FantasySCOTUS—towards the notion that judges of all ideological stripes are independent and decide cases solely based on the law.

¶115 These results are perhaps more honest, and sober, as players are voting their actual preferences, anonymously, with the incentive to win by accurately predicting votes. Were the same lawyers—excluding members of the professoriate, perhaps—to be polled formally, even anonymously, it is doubtful they would be so candid about their views of the Justices.[184] Rather, their answers may more likely be driven by platitudes as to what they think they should answer. Ultimately, the perceptions of Court watchers no doubt spill over, and affect the perceptions of the how the public at large views the Supreme Court. Determining an accurate picture of how Court watchers view the Court likely provides a window into how society at large views the Court.

¶116 These observations have several potential jurisprudential implications. FantasySCOTUS brings into stark focus that "[w]e are all realists now."[185] Predicting many cases, particularly the 5–4 splits, in a similar fashion to how people predict the outcome of political elections—during the last election, this district voted Republican, so it is likely to vote Republican again during the next election, regardless of the candidate's merits—reduces the judicial process from abstract, objective pronouncements of law to ascertaining the ideological votes of individual Justices. A survey of several power predictors suggested as much—members made predictions based on philosophical and ideological understandings of the Justices, sometimes without any regards to the merits of the actual case. As Prediction Markets grow more sophisticated, questions about the ideological Court, the nature of the judicial process, and the rule of law may become more pronounced.[186]

B. The Legal Prediction Market of Tomorrow

¶117 In a prescient 2005 article, Miriam Cherry and Robert Rogers postulated about an information market to predict Supreme Court decisions, named Tiresias, after the clairvoyant prophet of Thebes.[187] The authors remarked that "[t]he ability to know a

---

[184] A recent Pew Research Center Publication found that 39% of those surveyed view the Court as "middle of the road," 23% found the Court conservative, and 23% found the Court liberal. *The Invisible Court*, *supra* note 180. Overall, 58% of those surveyed have a favorable impression of the Court. *Id.* No data exist as to perceptions of each individual Justice. *Id.* Considering 53% of those surveyed could not name the Chief Justice of the United States, such data are likely impossible to collect among the public at large. *Id.*

[185] LAURA KALMAN, LEGAL REALISM AT YALE, 1927–1960, at 229 (1986).

[186] STEPHEN BREYER, MAKING OUR DEMOCRACY WORK: A JUDGE'S VIEW xiv (2010) ("At the end of the day, the public's confidence is what permits the Court to ensure a Constitution that is more than words on paper. It is what enables the Court to ensure that the Constitution functions democratically, that it protects individual liberty, and that it works in practice for the benefit of all Americans.").

[187] *See* Cherry & Rogers, *supra* note 46.

probable Supreme Court outcome in advance can potentially create monetary value for practitioners, provide guidance for lower courts, and perhaps even influence the Supreme Court itself."[188] Every year, the article notes, "probably hundreds, if not thousands, of civil disputes and criminal prosecutions are settled that contain issues the Supreme Court may resolve that Term."[189] Indeed, in light of the fact that the Court hears about eighty cases each year on a variety of topics, "many with monetary ramifications, the financial value of the Tiresias predictions could be considerable."[190] Over 100 years ago, Oliver Wendell Holmes, Jr. wrote, "[t]he object of our study, then, is prediction, the prediction of the incidence of the public force through the instrumentality of the courts."[191]

¶118 FantasySCOTUS takes a first step towards Tiresias, and fulfilling Holmes' observation, by creating a prediction market that could transform how attorneys make decisions. Future iterations of FantasySCOTUS will be more accurate, robust, powerful, and insightful. The software will be able to sense subtle changes in predictions at different stages of the litigation, and incorporate the historical performance of the Justices, and their voting patterns, along with the past success and track records of the power predictors automatically and instantly. In the future, the predictions will likely be accurate enough that people can meaningfully rely on them. Once the information market yields these rates, it could become an invaluable tool for litigation decisions.

¶119 Consider two cases—one civil, one criminal—recently decided by the Supreme Court, and how a FantasySCOTUS of the future, with a much higher accuracy rate, could provide helpful legal and litigation assistance for lawyers.

¶120 In *AT&T Mobility LLC v. Concepcion*, the Supreme Court found that California courts could not refuse to enforce contracts that prohibit class-action arbitration.[192] The case was argued on November 9, 2010 and decided on April 27, 2011.[193] Assume that before April 27, a Californian is threatening to assemble a class-action arbitration against a company, even though the contract the customer signed prohibits class-action arbitration. The company, following oral arguments in the lead-up to the Supreme Court's decision, is faced with a decision that could cost millions of dollars: risk a California court ordering costly class-action arbitration, or settle the matter and avoid the arbitration.

¶121 If the FantasySCOTUS of the future could predict with a degree of certainty that the Court will find that the contracts must be enforced, the company may be hesitant to engage in a settlement, as they will triumph in court. On the other hand, if FantasySCOTUS predicts the Supreme Court will agree with the California courts, and find the agreements unenforceable, the corporation may wish to settle the case, to avoid risky and expensive class-action arbitration. These are practical and tactical litigation decisions attorneys must make. Now, they can make this decision informed by data of

[188] *Id.* at 1142.
[189] *Id.* at 1183.
[190] *Id.*at 1184.
[191] Holmes, *supra* note 1, at 457.
[192] 131 S. Ct. 1740 (2011).
[193] *Id.*

what the Court will do, whereas in the past such decisions were made perhaps based on a law firm partner's "gut" instinct.[194]

¶122 The stakes in a criminal case could be even greater. Imagine that during an interrogation a suspect was read her *Miranda* rights, did not affirmatively invoke her right to remain silent, and subsequently made an incriminating statement. Assume that *Berghuis v. Thompkins*, which presented just this issue, has been argued before the Court, but not yet decided.[195] The prosecutor offers the defendant a favorable plea bargain that is only on the table for a limited duration; if not accepted, the prosecutor will take the case to trial. If the defendant accepts the plea agreement, she waives all appeal rights.

¶123 The defense attorney is faced with a choice. If her client accepts the plea bargain, and the Supreme Court subsequently finds that this interrogation did not result in a violation of *Miranda*, her client will have secured a short sentence, less than what she likely would have received at trial. Alternatively, if her client accepts the plea bargain, and the Supreme Court finds this interrogation did result in a violation of *Miranda*, her client cannot challenge the confession on appeal, and she is stuck in jail; had she gone to trial, the court would have suppressed the evidence, and she would have likely been acquitted without the confession.

¶124 If FantasySCOTUS shows that the Court will find a violation of *Miranda* rights in *Berghuis v. Thompkins*, perhaps the attorney should roll the dice and go to trial, hoping the judge will ultimately suppress the evidence, or perhaps her client could challenge it on appeal. If FantasySCOTUS shows that the Court will not find a violation of *Miranda* (the actual outcome of this 5–4 decision), perhaps the attorney should accept the favorable plea bargain, and not risk it. These are real decisions defense attorneys have to make. With the FantasySCOTUS of the future, this decision could be aided by informed predictions and their accompanying statistical measures of certainty.

C. From a Crowdsourced Prediction Market to an Intelligent Litigation Assistant

¶125 Admittedly, in its present form, FantasySCOTUS 1.0 is not particularly reliable for making important legal decisions. Further, while the eighty or so cases the Supreme Court decides each year are no doubt quite significant and of broad interest,[196] the 282,307 civil cases commenced in federal district courts[197] and the 56,790 appeals commenced in federal circuit courts in 2010 affect far more people.[198] A prediction market that can provide accurate predictions for the vast number of cases filed and appealed in federal courts each year could prove invaluable to lawyers and non-lawyers

[194] *See* AYRES, *supra* note 5, at 12 (With Super Cruncher information, "you don't need to guess, follow rules of thumb, or trust grizzled traditionalists. Increasingly, it is possible to tease out measureable effects of separate attributes to tell you what" approach would work best.).
[195] 130 S. Ct. 2250 (2010).
[196] *See* SUP. CT. R. 10.
[197] JAMES C. DUFF, ADMIN. OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS MARCH 31, 2010, at 40 tbl.C (2010), *available at* http://www.uscourts.gov/Viewer.aspx?doc=/uscourts/Statistics/FederalJudicialCaseloadStatistics/2010/tables/C00Mar10.pdf.
[198] *Id.* at 21 tbl.B, *available at* http://www.uscourts.gov/Viewer.aspx?doc=/uscourts/Statistics/FederalJudicialCaseloadStatistics/2010/tables/B00Mar10.pdf.

alike. Building on an idea developed by Professors Kobayashi and Ribstein in *Law's Information Revolution*,[199] a future version of FantasySCOTUS could shift from using a crowdsourced model (it is not likely that enough people will be intimately familiar with the thousands of cases decided in the inferior courts) to an algorithm that can Super Crunch data with an improved decision engine. The model would analyze data from previously decided cases to offer predictions for cases not yet filed.

¶126 It would be quite conceivable for a bot to crawl through all of the filings in PACER[200]—which stores every brief, opinion, and order filed in the federal courts, reportedly around 500 million documents[201]—and develop a comprehensive database of all aspects of how each court works. Using sophisticated text-recognition and natural language searches, a database could automatically index all of the cases—eliminating the need for fallible research assistants to laboriously tag cases. The system would note, for example, the parties to the case, the author and nature of a filed brief, the court it is filed with, the judge overseeing the case, the type of case it is, the damages or relief sought, the alleged merits of the case, the timeline of the case, the ultimate resolution of the case, and so on. This process would be instantly performed with every new filing, so the database would always be up to date with the latest jurisprudential and litigation trends, eliminating the need to resort to outdated data sets from the past.

¶127 With these data, a prediction engine could determine the various traits of successful and unsuccessful actions of various types, in various courts, under various circumstances. With enough data the prediction engine could provide, *ex ante*, a prognosis of how a case will likely proceed. Telling a client how a case will turn out—usually any client's main concern—is something that attorneys, no matter how well qualified, can only do imprecisely. As Professor Ayres remarked, "[t]rolling through databases can reveal underlying causes that traditional experts"—even pricey, experienced lawyers—"never even considered."[202] If lawyers could ascertain in advance what the likely results of litigation would be, they could "avoid[] disputes altogether"[203] and settle out of court. Even if the dispute cannot be avoided, a realistic prediction of probable damages could yield "ways to contain disagreements amicably and to avoid unnecessary escalation."[204]

¶128 But what if the engine could tell an attorney not only what will happen, but also how it should be accomplished? Imagine a program similar to the iPhone's Siri application. Call it *Harlan*. A would-be litigator could tell Harlan the relevant parties, the facts, the merits, and the remedy sought and share any relevant documents.[205] Harlan could generate a roadmap of how the case would be resolved with different judges in

[199] Kobayashi & Ribstein, *supra* note 9, at 1201 ("Lawyers might collaborate with computer scientists to develop new computer prediction algorithms," combing through public court records, such as PACER "to predict case results.").

[200] For a present-day tool that combs through PACER, consider RECAP, a crowdsourced program which allows people to "donate the documents they purchase from PACER" to "build a free and open repository of public court records." *About*, RECAP, https://www.recapthelaw.org/about/ (last visited Dec. 4, 2011).

[201] Timothy B. Lee, *Studying the Frequency in Redaction Failures in PACER*, FREEDOM TO TINKER (May 25, 2011, 1:52 PM), http://freedom-to-tinker.com/blog/tblee/studying-frequency-redaction-failures-pacer.

[202] AYRES, *supra* note 5, at 12.

[203] RICHARD SUSSKIND, THE END OF LAWYERS? RETHINKING THE NATURE OF LEGAL SERVICES 184 (2008).

[204] *Id.*

[205] For an example of how a Harlan simulation could provide litigation assistance to lawyers and non-lawyers alike, see Blackman, *supra* note 8.

different courts, and perhaps even recommend an ideal forum (call it fantasy-forum-shopping). Harlan could explain how best to structure the litigation, what types of motions would be most successful, and how to arrange arguments. With advances in artificial intelligence—Google developed cars that drive themselves, and IBM's Watson defeated the Jeopardy world champion[206]—it is not much of a stretch to suggest that Harlan could even draft the briefs (many sections of briefs today are copied from boilerplate anyway), or at least check the persuasiveness of the arguments against other successful arguments already accepted by courts. Harlan would also work wonders for non-lawyers. A person could download the app, talk to *Harlan* in plain language, explain the problem, and listen to possible remedies—that may or may not involve paying a lawyer. Harlan would improve access to justice, at little to no cost.

¶129 Such a product would transform the legal profession and our society. This change would require a fundamental rethinking of approaches to legal education,[207] the practice of law, and, broadly speaking, our system of justice. It will likely first be first met with doubt—"computers can't replace human lawyers!" This technology would not be about replacing lawyers (at least not lawyers who adapt[208]); rather, it would provide advocates with information and knowledge to serve clients more effectively and at a lower cost. Next, there will likely be fierce resistance to change from entrenched interests in the form of ethical and regulatory challenges[209]—"computers can't follow the rules of ethics and they will provide ineffective legal assistance to non-lawyers!" These criticisms are fair, but such technology could provide opportunities to improve the quality of representation to all segments of society. Rather than instinctively opposing any change that upsets the status quo, these new technologies should be met with tempered enthusiasm. Reforms to the regulatory regime will come,[210] followed by gradual acceptance of this technology. We hope that FantasySCOTUS will serve as a first step in the evolution from today's time-consuming, customized labor-intensive legal market to tomorrow's on-demand, commoditized law's information revolution.[211]

---

[206] ERIK BRYNJOLFSSON & ANDREW MCAFEE, RACE AGAINST THE MACHINE: HOW THE DIGITAL REVOLUTION IS ACCELERATING INNOVATION, DRIVING PRODUCTIVITY, AND IRREVERSIBLY TRANSFORMING EMPLOYMENT AND THE ECONOMY 316–18 (2011) ("[L]ike the Google autonomous car, Watson the *Jeopardy!* champion supercomputer, and high-quality instantaneous machine translation, then, can be seen as the first examples of the kinds of digital innovations we'll see as we move further into the second half—into the phase where exponential growth yields jaw-dropping results.").

[207] *See* Larry E. Ribstein, *Practicing Theory: Legal Education for the Twenty-first Century*, 96 IOWA L. REV. 1649 (2011).

[208] *See* Blackman, *supra* note 8.

[209] Suits against LegalZoom.com for the unauthorized practice of law—a nebulous and vague term that will likely have to be defined by courts to reflect the development of the legal profession—are the first glimpses of the future of this litigation. Debra Cassens Weiss, *LegalZoom Can Continue to Offer Documents in Missouri Under Proposed Settlement*, A.B.A. J. NEWS (Aug. 23, 2011, 6:32 AM), http://www.abajournal.com/news/article/legalzoom_can_continue_to_offer_documents_in_missouri_under_proposed_settle/; Debra Cassens Weiss, *LegalZoom Sues North Carolina State Bar, Seeks to Register Legal Services Plan*, A.B.A. J. NEWS (Oct. 5, 2011, 5:31 AM), http://www.abajournal.com/news/article/legalzoom_sues_north_carolina_state_bar_seeks_to_register_legal_services_pl/.

[210] *See generally Unlocking the Law Symposium*, TRUTH ON MARKET, http://truthonthemarket.com/unlocking-the-law-symposium/ (last visited Dec. 29, 2011).

[211] *See* BRYNJOLFSSON & MCAFEE, *supra* note 206, at 363; MICHIO KAKU, PHYSICS OF THE FUTURE: HOW SCIENCE WILL SHAPE HUMAN DESTINY AND OUR DAILY LIVES BY THE YEAR 2100, at 312–13 (2011)

VIII. CONCLUSION

¶130 The inner-workings of the Supreme Court of the United States are shrouded in secrecy. From the first Monday in October until the last week in June, the Justices operate behind-the-scenes to decide some of the most important issues in our society. Now FantasySCOTUS can provide real-time predictions how the Court will decide these cases. The FantasySCOTUS crowdsourced prediction market provides a novel insight into how Court watchers perceive the decision-making of the United States Supreme Court.

This essay lays the foundation for future research into the predictive power of FantasySCOTUS. Ultimately, the data that serve as the basis for this Article are simply a starting point. As FantasySCOTUS continues to crowdsource new information, we will gain new and deeper insights into the task of predicting Supreme Court cases and modeling judicial behavior. Looking forward, this project is not just a scholarly exposition of a theoretical construct or a discussion of a novel fantasy league that yields respectable, but an analysis of not-yet reliable, Supreme Court predictions. Rather, it is effectively an emerging plan for a legal information service that could transform the way lawyers, and non-lawyers alike, interact with courts in the not-so-distant future.

(“When technologies become widely dispersed, such as electricity and running water, they eventually become utilities. With capitalism driving down prices and increasing competition, these technologies will be sold like utilities, that is, we don’t care where they come from and we pay for them only when we want them.”); SUSSKIND, *supra* note 203, at 32 (“In summary, a commoditized legal service is an IT-based offering that is undifferentiated in the marketplace (undifferentiated in the minds of the recipients and not the providers of the service). For any given commodity, there may be very similar competitor products, or the product is so commonplace that it is distributed at low or no cost.”); Kobayashi & Ribstein, *supra* note 9, at 1218 (“[T]he opportunities evident in advances in information technology will make more visible the costs of maintaining the current system of relying on the one-to-one delivery of legal advice and the benefits of moving to a legal information market.”).