In the internet age, teams are accomplishing complex tasks in management. However, relatively little attention has been directed towards optimising work group processes to reduce random or systematic error (bias) in order to improve performance and optimise decision making. Internet-based prediction markets have been developed to meet this need. Prediction markets can be used for forecasting and the performance of responders can be assessed when the future event actually occurs. Combinatorial prediction markets forecast the likelihood of combined events.1

Here, we aim to compare traditional open group work to internet-enhanced crowdsourcing and, in particular, prediction markets. We describe prediction markets and how their performance compares to other sources of information gathering and decision making. We describe the current and potential uses of prediction markets for governance within organisations, and the related advantages and disadvantages.

Questions that could be opened to a prediction market in a financial institution include:

- **Accounting:** ‘What will be the loan loss provision for the department for this half-year?’
- **Audit:** ‘At next review, what will the severity code be on issue ABC identified in this year’s audit?’
- **Compliance:** ‘What new compliance breach will be flagged in this quarter’s review?’
- **Credit:** ‘At next review, what will be the risk grade on customer Z?’
- **Markets:** ‘What will competitor Bank Beta’s RBA rate cut pass-through be on their standard mortgage rate?’
- **Project management:** ‘What will the red/amber/green status be on project alpha at the end of next month?’
- **Sales:** ‘Will customer 123 close the deal this quarter?’

Crowdsourcing platforms can enhance an organisation’s management decisions and governance, by harnessing the ‘wisdom of the crowds’. Prediction markets go one step further: to also provide an iterative summary signal of the crowd estimate back to participants. Evidence is accumulating that prediction markets can perform better than opinion experts, management consultants and surveys under specific conditions. Multinational and other companies are currently using prediction markets predominantly for improved information delivery, and adopting reward structures to induce informed participation by employees. This is an abridged and revised version of a paper presented at the 2016 Melbourne Money and Finance Conference.

KARL MATTINGLY, CEO, affiliated with Almanis, Percrypt and Dysrupt Pty Ltd, Melbourne
ANNE-LOUISE PONSONBY, Professor of Epidemiology, affiliated with Population Health, Murdoch Childrens Research Institute, Royal Children’s Hospital Melbourne, School of Population and Global Health, University of Melbourne and Member, Dysrupt Advisory Board
How crowdsourcing platforms such as prediction markets add value beyond open group work

Our familiar group process in business is that of open group work. Traditionally, this has occurred face to face, but now is often performed electronically. It provides real-time multidisciplinary input and interactive feedback but can have several disadvantages.

First, for effective communication, open group size needs to be limited and Wheelan et al. (2009) found that groups of three to eight were generally more productive than larger groups. If intragroup communication is not required, very large teams will have advantages of greater human resource capital and larger cognitive diversity. Crowdsourcing is then ideal.2

Second, intragroup social influence can reduce the diversity of individual opinions without reducing collective error.3 Individual biases may become correlated and a new set of social biases introduced at the interactive group level (Table 1).

It is well established that independent aggregation of multiple independent opinions through simple averaging, majority rules or market-based algorithms have led to a marked improvement in decision accuracy. This phenomenon is known as the ‘wisdom of the crowd’ effect.4 Not only is the group aggregate average decision accurate but only a very small minority of individuals can perform consistently better than the group average.5 An advantage of a collective group aggregate approach is that the group size can be very large, maximising cognitive diversity without social bias. Page (2007) provides evidence that even as the average amount of expertise decreases when a crowd grows, the increased diversity may more than compensate for this.

Crowdsourcing social software platforms have emerged to service the need for big data and associated analytics, whether through internet platforms like Survey Monkey, or through prize-based competitions. Prediction markets provide further enhancement by providing a continually updated group aggregate summary estimate as an iterative feedback signal to the wider group. This signal can take the form of an updated summary of the probability of event occurrence or the likely value. This important concept has been taken up for market-based decisions on outcomes. In this collective intelligence process no direct member-member communication is required.

Crowdsourcing platforms harness the ‘wisdom of the crowds’, which is partly derived from the greater cognitive diversity and chance of uncovering hidden information from very large groups and partly due to a lack of social bias if individuals within the crowd provide independent input. Prediction markets go one step further: to also provide an iterative summary signal of the crowd estimate back to participants in the crowdsourcing platform.

Individualising incentives for participants in crowdsourced prediction markets

Because the prediction markets detect both effort and individual informative and accurate input, standard market mechanisms can be used to deliver rewards providing incentives that are based not only on activity but on forecasting accuracy and informativeness. To reward based on the informativeness of the participants’ trades, one must judge the quality of participants’ forecasts once the outcome is known. This enables participants to be rewarded so that they are incentivised to continue forecasting. Market algorithms, such as Hanson’s logarithmic market scoring rule (LMSR),6 provide an effective method of assessing the quality of a participant’s forecast on a single question by evaluating the amount of additional information the particular forecast provided to the group’s aggregate forecast on the question.

Reputational incentives have also been demonstrated to be important. These include Leaderboards, listing high-performing participants (often by a codename) or badges. Prediction markets can adopt some of the real time psychic utility of ‘gamification’.7 Finally, altruistic and charitable motivations can be included. Mueller (2008) observes that in a large open work group a lack of individual reward or recognition linked to personal effort can lead to social loafing. In the crowdsourcing platform of a prediction market, these incentives — monetary, reputational, experiential and altruistic — can be targeted back to the individual based on the merit of their specific contribution.
Table 1: Key biases that operate in open groups

- **Group think**: desire for harmony or conformity within the group decreases external input and critical thinking, results in an incorrect decision-making outcome. Group members try to minimize conflict and reach a consensus decision without critical evaluation of alternative ideas or viewpoints, and by isolating themselves from outside influences.

- **Halo effect**: the tendency for a person's positive or negative traits to ‘spill over’ from one area of their persona to another in others' perceptions of them. Thus social or status cues can weight the relative importance of an individual's input, regardless of actual merit.

- **Unacceptability bias**: questions that may embarrass or invade privacy are avoided and hidden information is not disclosed.

Standard market mechanisms can be used in a prediction market to deliver incentive-based monetary rewards to individuals not only based on activity but also forecasting accuracy and informativeness.

**Potential problems with prediction markets**

First, there is an *inefficiency problem*, due to lack of cognitive diversity when the market is small. For example, Graefe (2011) evaluates prediction markets compared to other group work processes and found no evidence of outperformance by prediction markets in small groups (of less than 10 members). Evidence of optimal prediction market size for various purposes is not yet available but it is likely that at least 30 to 50 members would be required to allow sufficient cognitive diversity. For thin markets, one advance to assist low liquidity markets is the LMSR.3

Second, prediction markets may not include *sufficient expertise*. However, prediction market membership rules can be altered to ameliorate this. Cowgill and Zitzewitz (2015) examine prediction markets used by Ford which only allowed staff with expertise to participate.

Third, prediction markets could be susceptible to *manipulation*. Yet, to date, prediction markets appear relatively resistant with only small and temporary effects on trading. Also, difficult questions may be more prone to *misinterpretation* than by face-to-face query. However, complex questions such as the effect of a possible rare event on an outcome can be broken down into simpler components, such as: (a) what is the probability of the event occurring?; and (b), if it does, what is the impact on the outcome? In a bank credit assessment situation, the prediction market could ask for separate estimates of the probability of default (within some given timeframe) of a specified borrower, and of the size of the loss given default, rather than the expected loss from granting a loan to that borrower.1

Finally, one obstacle to the uptake of prediction markets within small businesses and other organisations is the *complexity of development and use*. To date, most companies (Google being a major exception), have required external consulting expertise to build and operate in-house prediction markets. Dysrupt labs (with which the authors are associated) has developed a user-configurable platform with a simple user interface and supporting use case collaterals that dramatically reduce the initialisation costs for a prediction market. This enables low cost deployment for simple recurring forecasts as well as the more complex bespoke problems and question sets that may be expensive to initialise in a traditional prediction market.

**Evaluation studies of prediction market performance**

Prediction markets have been found to outperform expert panels, professional management consultants and simple or competence based-surveys and polls in the following ways.

*Official estimates and experts*: Plott and Chen (2002) report that at Hewlett Packard, prediction markets predicted outcomes six out of eight times more reliably than official forecasts in one experiment and 15 out of 16 in another. Eli Lily, a pharmaceutical company, found a prediction market was able to prospectively identify the three most popular sales products. Siemens reports that an in-house prediction market provided a more reliable project end date than official management estimates. SciCast reports that their prediction market generally outperformed an unweighted linear opinion poll of experts seven out of 10 times across a range of issues, including geopolitical. Higher performance for prediction markets than expert forecasting was documented for the three case studies of in-house prediction markets, reviewed by Cowgill and Zitzewitz (2015).
Consultants: Thompson (2012) reports that in 2007 the CEO of Misys ran an in-house prediction market in parallel with a McKinsey management consulting assignment including interviewing staff to determine preferred strategy over the next three years. The questions for strategy were similar for the McKinsey consulting team and the prediction market. They both gave the same answer but the prediction market was much more cost-effective because it could be then used repeatedly once purchased and the initial purchase price was similar to a one-off McKinsey consult. Subsequently, Thompson reports Misys prediction markets proved to be 96 per cent accurate measured against actual outcomes, much more accurate than the company’s own internal three-month forecasts, compiled using usual business processes.

Surveys: Surveys face some shortcomings that make them less appropriate tools than prediction markets for making decisions on complex issues. There is little incentive for a survey respondent to give their best answers, as rewards tend to be based purely on participation rather than the quality of responses. In addition, once an early respondent provides an answer that is a fixed answer, there are generally no ‘feedback loops’ to guide them to re-evaluate their decision. These may be some of the reasons underlying the accumulating evidence that surveys do not forecast as well as prediction markets.

Perhaps some of the strongest evidence to date comes from the evaluation of the reproducibility of results of over 40 studies published in prominent psychology journals. Prediction markets forecasted the outcomes of future replication studies well and outperformed a survey of individual forecasts. Compared to a reference point of correctly identifying 50 per cent of replication studies by chance, the prediction markets correctly predicted the outcome of 71 per cent of scientific replication studies, significantly more than expected. In comparison, an average of a simple survey predicted 58 per cent, and an average of a survey weighted by self-reported expertise predicted 50 per cent.\(^8\) As further evidence, on the SciCast platform, a prediction market performed 20 per cent better than an unweighted baseline survey.

Polls: The public prediction market Iowa Electronic Markets performance consistently outpredicted opinion polls for election results from 1988 to 2004. It is one of the earliest prediction markets and has significantly outperformed the polls in every presidential election when forecasting more than 100 days in advance: Berg (2008) compares 964 polls over the five presidential elections from 1988 to 2004: the Iowa market was closer than the poll to the eventual outcome 74 per cent of the time.

Prediction markets: These do differ, but not substantially, in their performance, provided that the respondent number is large and the question is appropriately framed. There are numerous ways to measure the accuracy of a forecasting prediction market platform, but the most common method is the Brier Score, which is in effect the mean squared error of the forecasted likelihoods. A perfect forecasting platform will receive a score of 0. A platform that forecasts no better than chance will receive a score of 0.5. Perhaps a more intuitive measure of a platform’s accuracy is how often the forecast is on the right side of Yes/No. While a forecasting platform’s accuracy will generally be best just before the answer is objectively known, knowing likelihoods at this point is not always useful; it is often too late to take any action utilising the data. From November 2015 to March 2016, the Almanis prediction market was on the right side of Yes/No 70 per cent of the time for 10 days before settlement, improving to 90 per cent of the time at 0 days before settlement. Topics covered included geopolitics, economics and finance. Meanwhile, the Brier Score ranges from around 0.4 at 10 days out, down to 0.2 at 0 days. The Brier scores achieved by Almanis compare similarly well to other prediction markets with accessible Brier scores when predicting US Senate Election contests.\(^9\)
Prediction markets, information flow, and governance in organisations

Figure 1 outlines a traditional corporate structure with work specialised into context teams such as finance, human resources, marketing, production and other. Each team is led by a chief, such as chief financial officer, who sits in the executive suite, the C-suite, with the chief executive officer. In this example, the CFO and the CEO sit on the Board of Directors also.

Figure 1: Actual and potential functions for prediction markets in corporations

Abramowicz and Henderson (2006) review the ways that prediction markets could enhance the corporate governance of traditional organisations. Traditional organisations are characterised by hierarchies, where those at a lower level act as a synthesiser, summariser and gatekeeper for the collective information flow upwards (Figure 1). To some extent, this assists in the reduction of an overwhelming data load. However, self-interested actors at each level can distort this process and the information from hidden experts may not be heard, reducing transparency. Recent corporate responses include monitoring and whistle-blower protection. A key role for prediction markets,
which is already utilised by a number of companies, is the improvement of information delivery
to senior executives charged with corporate decision making. In Figure 1, this is shown within a
content team (such as the finance function generating information for the CFO) and across an
entire organisation generating information for the C-suite. If in-house staff trade anonymously
and are given incentives to provide informed and unbiased information, the signal from those
in-house who are knowledgeable becomes strong, and this is also enhanced by the market
signal feedback loop. This in-house prediction market does not replace but supplements existing
practices, as has been the case with the companies that have employed such markets.

Further, prediction markets can assist senior executive decision making. The literature indicates
that supplementary information and checks improve management decision quality. An in-house
prediction market can be used to confirm or corroborate management’s opinion. Unfortunately,
the size of most boards precludes use within the board membership only.

Shareholder signals act through a public share price signal from the market of the company’s
shares and informs the Board. However, this does not provide detailed information on
shareholder views on various different aspects of the specific company. Public share prices have
some relative disadvantages as they do not include in-house information, can only respond to
one factor at a time (not alternatives) and may reflect multiple other external factors such as
politics rather than company performance alone. A shareholder prediction market could be run
on the merit or likelihood of possible mergers, asset sales, takeover defenses or relocation to
another region (Figure 1). For high-value issues, the Board could run parallel prediction markets
among both shareholders and staff to evaluate their opinions on the effect on productivity
or various new scenarios, and the latter could be used with the benefit of anonymously
incorporating any insider information.

Auditing is another governance function requiring accurate information elicitation and benefiting
from forecasting. Here, prediction markets could play a valuable role in the internal audit, external
audit and regulation of standardised audit processes. Within a content team such as finance,
the CFO could run a prediction market as a pre-diagnostic before an official external audit to
evaluate issues such as whether the group will be more or less likely to score better than the
previous year in specific areas and identify new areas of concern. Similar processes apply in using
a prediction market preceding an external audit (Figure 1). Finally, Abramowicz and Henderson
(2006) state that prediction markets may also have value as a substitute for (or complement
to) existing securities disclosure requirements under regulatory law. Regulatory and legislator
bodies could set up standardised routine prediction markets to evaluate aspects of company
performance by mechanisms such as external audit (Figure 1).

Thus we can see that prediction market use can improve transparency within an organisation
in several ways, as well as improving the quality of information to be used by the C suite for
decision making. The C-suite, in turn, may be subject to greater accountability to their Boards
and the market. For example, where an in-house prediction market is at odds with management
representations on an issue of governance (risk, audit or compliance) it would be important
to reconcile the points of difference rather than censor the results in order to manufacture
the illusion of consensus [or plausible deniability] for the Board. Similarly, where censorship
and selective use of prediction market accuracy is pervasive, the effect is to degrade market
participation and thus the quality of the market’s signal.

Company use of prediction markets with case examples

Google has run in-house prediction markets since 2005. Some other companies, without the
internal resources to develop prediction markets, have been able to work with professional
research and consulting firms to develop appropriate in-house prediction markets but this has
been beyond the financial and technical reach of most small businesses and other organisations.
For example, at the time of journal submission, an internet search was unable to identify any
prediction markets in companies based in Australia or New Zealand, apart from the Almanis and
Percrypt markets within Dysrupt Pty Ltd.

Cowgill and Zitzewitz (2015) provide a detailed review of prediction market performance over
selected periods for three major companies: Google, Ford Motor Company and Firm X. They
also list a number of other companies running similar markets including Chrysler, Eli Lily, GE,
Best Buy, Boeing and Microsoft. We draw on their study to provide some salient features of the
characteristics of those markets.
The sorts of questions posed in the prediction markets included company performance, demand forecasting, project completion dates, product quality and external events. To encourage employee participation a range of incentives have been offered including monetary prizes, non-monetary prizes (such as T-shirts) and reputational recognition.

Google, a world dominant software company based in California, has a very educated workforce and an organisational culture of high transparency. Google’s prediction markets were initiated by a group of employees and were of the continuous double auction form, based on the style of the Iowa Election Markets. All employees could participate. One question example was whether chat would be launched by Gmail by the end of the quarter.

Ford Motor Company chose to focus its prediction markets on two topics: forecasting weekly sales volumes; and predicting which car features would be popular with customers (as proxied in the interim by traditional market research, including focus groups or surveys). Firm X is a global private diversified basic materials and energy organisation headquartered in Midwestern US. It decided to focus its prediction markets on relevant macroeconomic and commodity prices. Firm X’s markets were started by a senior manager in its strategic planning department. For Ford and Firm X, the LMSR was applied as the central market mechanism and only employees with expertise were invited to participate. Overall, despite some possible limitations partly due to Firm X sample size, the prediction markets performed better than forecasts by experts.

Today, multinational and other companies are currently using prediction markets predominantly for improved information delivery to senior executives charged with corporate decision making. This ranges from higher forecasting accuracy of the probability of external events, sales or alterations in company product value to in-house ideas generation.

Conclusion
We have compared group open work to internet-enhanced crowdsourcing and, in particular, prediction markets as a tool for channeling collective intelligence. We outlined the performance of prediction markets compared to other methods such as surveys and related their use within traditional corporate governance. To date, prediction markets have mainly been used to improve information delivery but their potential role in business is likely to grow.

Acknowledgements
We thank Jane Carroll, Alex Gruen and Stephanie Ponsonby for literature review and manuscript preparation. We thank Sam Dentry and Nick Williams for information on Almanis metrics. Anne-Louise Ponsonby holds a Senior National Medical Research Fellowship. This work was supported by the Australian Government Research & Development Tax Incentive Scheme and a Victorian Government Infrastructure Operational Grant.

Notes
1. Such a combined probability is called a conditional or joint probability (Ryall and Bramson 2003). Combinatorial prediction markets have been designed (Hanson 2007) but they are not common.
2. Excellent reviews, including historical aspects, of the ‘wisdom of the crowds’ and ‘collective wisdom’ are provided by Surowiecki (2004) and Landemore (2012), respectively.
6. In decision theory, a scoring rule measures the accuracy of probabilistic predictions. It is applicable to tasks in which predictions must assign probabilities to a set of mutually exclusive discrete outcomes. The logarithmic market-scoring rule can be run as an automated market algorithm, even in relatively low liquidity markets (Hanson 2004, 2007).
7. Considerable useful psychological research has emerged from studies of online gaming that may provide useful insights for other human internet use. For example, gamers can be characterised into the extremes of social players motivated by relationships and teamwork and achievement players valuing progress, optimisation and domination. Dr Rosanna Guadagno University of Texas at Dallas. Presentation at the 16th annual meeting of the Association of Internet Researchers in Phoenix, Texas, US, 2016.

8. A linear probability model with robust standard errors with the outcome of the replication study as the independent variable and the prediction market price as the dependent variable found the beta coefficient of market price to be 0.995, which is significantly different ($p=0.003$) to 0 (no predictive effect) and not significantly different ($p=0.99$) to 1 (a very high predictive effect), see Dreber (2016).

9. From November 2015 to November 2016, on the Almanis prediction market platform, the accuracy of questions by their content domains were ranked as follows: geopolitical events (highest), medical, light relief, social, technology, industry, economics, markets (lowest).


11. Available at http://www.almanis.com/

12. In all four evaluations, the prediction market forecast had a higher predictive weight with a lower mean squared error than the expert forecast. In three of four, the difference was $p<0.01$. The Ford prediction market-expert forecast comparison was not significantly different; $p=0.10$. It should be noted that Ford was the lowest volume evaluation: 6 (vs. 191–296) unique markets; 78 (vs. 197–296) observations and this reduced sample, with lower statistical power, may have partly contributed to this finding. Nevertheless, the magnitude of the mean-squared error improvement achieved by the Ford prediction market was the largest, see Cowgill and Zitzewitz (2015).

References


Armstrong, JS 2006, ‘How to make better forecasts and decisions: Avoid face-to-face meetings,’ Foresight — The International Journal of Applied Forecasting, Fall, no. 5, p. 12.


Bratvold, D 2011, Enterprise Crowd sourcing blasts off as social media growth industry, MarkSchafer.


Rhode, P and Strumpf, K 2007, ‘Manipulating political stock markets: A field experiment and a century of observational data’, working paper, University of Arizona and NBER, University of Kansas School of Business.


Survey Monkey.


