Accuracy-Risk Trade-Off Due to Social Learning in Crowd-Sourced Financial Predictions

Dhaval Adjodah [1,*] **, Yan Leng** [2] **, Shi Kai Chong** [1] **, P. M. Krafft** [3] **, Esteban Moro** [4] **and Alex Pentland** [1,*]

1   Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; cshikai@mit.edu
2   McCombs School of Business, The University of Texas at Austin, Austin, TX 78712, USA; yleng@mit.edu
3   Oxford Internet Institute, University of Oxford, Oxford OX1 2JD, UK; pkrafft@mit.edu
4   Departamento de Matemáticas & GISC, Universidad Carlos III de Madrid, 28911 Leganes, Spain; emoro@mit.edu
*   Correspondence: dval@mit.edu (D.A.); pentland@mit.edu (A.P.)

Abstract: A critical question relevant to the increasing importance of crowd-sourced-based finance is how to optimize collective information processing and decision-making. Here, we investigate an often under-studied aspect of the performance of online traders: beyond focusing on just accuracy, what gives rise to the trade-off between risk and accuracy at the collective level? Answers to this question will lead to designing and deploying more effective crowd-sourced financial platforms and to minimizing issues stemming from risk such as implied volatility. To investigate this trade-off, we conducted a large online Wisdom of the Crowd study where 2037 participants predicted the prices of real financial assets (S&P 500, WTI Oil and Gold prices). Using the data collected, we modeled the belief update process of participants using models inspired by Bayesian models of cognition. We show that subsets of predictions chosen based on their belief update strategies lie on a Pareto frontier between accuracy and risk, mediated by social learning. We also observe that social learning led to superior accuracy during one of our rounds that occurred during the high market uncertainty of the Brexit vote.

Keywords: crowd-sourcing; wisdom of the crowd; social learning; Bayesian models; risk

1. Introduction

Distributed financial platforms are on the rise, ranging from Decentralized Autonomous Organizations [1], crowd-sourced prediction systems [2] to the very recent events during which retail investors self-organized using social media and drove up asset and derivative prices [3,4]. In this work, we investigate how financial agents process information from one another and predict-individually and collectively—the future prices of real assets. Specifically, we are interested in understanding the computational models they use to update their beliefs after information exposure and how different social vs. non-social belief update strategies lead to trade-offs in prediction performance.

Here, we expand the typical definition of performance for collective prediction to include the concept of risk. Typically, the prediction performance of collectives and swarms is measured mostly by the accuracy of the group over collections of tasks [5–7]. However, it has been shown theoretically [8,9] and observed in a variety of applications [10,11] that there is a fundamental trade-off between prediction accuracy (average error) and prediction risk (variance of error).

This means that for any prediction system, risk will always be present, and that maximizing accuracy will come at the expense of increased risk. Hence, the performance of the system will always exist within a pre-defined Pareto frontier [12,13] which is the curve containing all possible system performance parametrizations (here, pairs of possible accuracy and risk values). Therefore, a platform designer will need to make trade-offs between risk and accuracy and cannot achieve arbitrarily combinations of risk and accuracy. Treating risk and accuracy as equally important for prediction is standard in

statistical [8–10] and financial [14–16] forecasting applications and literature because it allows for prediction systems to be calibrated and deployed with regard to specific accuracy and risk profiles [17–21].

However, characterizing the performance of crowd-based prediction systems regarding both accuracy and risk is not common and such a Pareto frontier has not been observed in crowd-sourced financial asset price prediction. We are therefore interested in investigating if a Pareto frontier exists and what the causes are behind this trade-off. From the perspective of crowd-sourced financial platform designers, understanding the trade-off between accuracy and risk and how to select subsets of predictions that achieve a certain accuracy and risk is useful to fit a required risk profile. This, in turn, allows for more sophisticated and versatile applications of crowd-sourced predictions such as hedging risks over portfolios of prediction tasks.

To test our hypothesis that a Pareto frontier exists between risk and accuracy and that it is mediated by social learning, we designed our collective prediction experiments as a series of Wisdom of the Crowd (WoC) tasks. For background, the Wisdom of the Crowd [22,23] is a popular domain within the collective intelligence literature where participants (the 'crowd') are asked to make predictions of a certain quantity, such as the future price of an asset on the stock market [24] or the caloric content of food items [25]. Prior work in the WoC literature [25–27] has focused on maximizing the average accuracy of collectives with little regard to the risk of the predictions.

The structure of this paper is as follows: we do a short literature review of the connections of this work to research on collective intelligence and the accuracy-risk trade-off in Section 2. We discuss our materials and methods (experimental design, data collection, and modeling and estimation) in Section 3. We present our results (belief update modeling, accuracy-risk trade-off and prediction under high uncertainty during Brexit) in Section 4. We discuss the implications and limitations of our work in Section 5.

Contributions

Our work makes the following novel contributions:

- We present an experimental procedure where we exposed 2037 participants to social and non-social information during 7 independent rounds of predicting financial asset prices (S&P 500, gold and WTI Oil). We collected 4634 prediction sets which include participants' predictions before and after information exposure, as well as the information they were exposed to. We are releasing this data here.
- Using computational models inspired by Bayesian models of cognition [28,29] to investigate the belief update strategy of participants, we observe that a simple model that approximates the likelihood (evidence) to be a unimodal Gaussian beats a more complex Monte Carlo approach. This suggests that our participants exhibit the attribute substitution heuristic of human decision-making [30], whereby a complicated problem is solved by approximating it with a simpler, less accurate model.
- We observe that participants prefer to learn from social information rather than from non-social information, another interesting information processing heuristic.
- Our main contribution: we observe a Pareto frontier between accuracy and risk. As the average accuracy of the crowd over the different prediction rounds increases, so does the risk in the crowd's predictive accuracy. We further observe that this trade-off is mediated by the amount of social learning i.e., the extent to which participants pay attention to each other's judgments.
- We deployed one of our prediction tasks just before the Brexit vote during which there was a great deal of market uncertainty [31], and we observe that during such uncertain times social learning leads to higher accuracy.

These results are not only important for the practical deployment of distributed financial prediction platforms but also expand our understanding of how financial agents process information and make distributed predictions.

2. Related Work

2.1. Collective Intelligence and Social Learning

There is a rich literature on how decentralized information processing, learning and decision-making affects the performance of collectives and swarms [32–36]. Here, we focus on how platforms can be designed for people to make predictions with high performance, which is a central question for the Wisdom of the Crowd [22,23,37].

It has been shown that the temporal influence and mutual information dynamics between individuals can have a strong effect on crowd collective performance. On the one hand, prior work has shown that exposure to social information can lead to degraded performance in aggregate guesses [26,37,38]. For example, increasing the strength of social influence has been shown to increase inequality [39]. Selecting the predictions of people who are resistant to social influence has been shown to have improved collective accuracy [27]. The influence of influential peers has been theoretically shown to prevent the group from converging on the true estimate [26], and exposure to the confidence levels of others has been shown to influence people to change their predictions for the worse [40].

On the other hand, social learning has also been shown to lead to groups outperforming their best individuals when they work separately [41] and a collective intelligence factor has been shown to predict team performance better than the maximum intelligence of members of the team [35]. Similarly, human-inspired social communication between agents has been shown to improve collective performance in optimization algorithms [5,42].

Therefore, the role of social learning in collective performance is still being understood. Our contribution to this line of research is that a more complete characterization of performance in terms of not just accuracy but also risk provides avenues for future work towards reconciling the disagreements as to the role of social influence on performance. This is especially important due to the already existing strong social components in many crowd-sourcing platforms and applications [43–48] that could be harnessed more effectively for performance improvement.

2.2. Accuracy-Risk Trade-Off

Previous work has investigated several avenues to optimize the accuracy of the crowd such as by recalibrating predictions against systematic biases of individuals [26] and selecting participants who are resistant to social influence [27]. Additionally, rewiring the network topology of information-sharing between subjects [25,41], and optimally allocating tasks to individuals [49] has improved collective accuracy. However, these studies focused on accuracy with little regard to risk. There is a rising movement to go beyond accuracy and to fully characterize performance—at the individual and the collective level—in terms of both accuracy and risk. Some call this emerging line of work going beyond the 'bias bias (In the statistics literature, bias is another name for accuracy. This movement suggests that research should go beyond its current focus on just bias and study risk).

At the individual level, there is increasing evidence that people preferentially optimize for risk instead of accuracy in a variety of domains [50]. Cognitively, people have been observed to manifest decision heuristics [51] to be conservative in the face of uncertainty [52,53]. For example, rice farmers have been observed not to adopt significant harvest improvement technology because of the risk of it failing once and causing significant family ruin [54]. Evolutionarily, risk aversion has been shown to emerge when rare events have a large impact on individual fitness [52]. Furthermore, in a meta-study of 105 forecasting papers, 102 of them support prioritizing for lower risk to achieve higher overall performance [55]. At the collective level, there is limited work regarding the characterization of the performance of collectives and swarms in terms of both accuracy and risk although there is a large literature on other related trade-offs such as between speed and accuracy [56–60].

From a system design perspective, crowd-sourcing platform designers should characterize their performance in terms of both accuracy and risk due to theoretical results [8,9] and observations in applications [10,11] that the performance of any prediction system is

subject to a fundamental trade-off between accuracy and risk. This is especially important in our domain of predicting financial asset prices as risk is already known to have negative effects on the efficiency of markets such as through the phenomenon of implied volatility [61].

3. Materials and Methods

3.1. Experimental Design

To test our hypothesis that a Pareto frontier exists between risk and accuracy—i.e., that there is a trade-off between risk and accuracy of prediction across several prediction rounds—and that it is mediated by social learning, we need a dataset with the following requirements:

- Predictions are made of complex and difficult-to-predict phenomena so that our results are applicable to the real-world platform applications.
- Predictions are made over many independent prediction rounds so that the risk of the crowd over these different tasks can be estimated.
- A ground-truth is needed against which we can compare our dataset to judge the external validity of individual and collective performance metric.
- The social and non-social information each participant was exposed to after their initial pre-exposure prediction is recorded so that we can later model how different types of information influenced them in updating their belief into their post-exposure prediction.

Given the above requirements, we designed the experimental procedure as detailed below: we recruited a total of 2037 participants over seven prediction rounds to predict the future prices of financial assets (the S&P 500, WTI Oil, and gold prices) during seven separate consecutive 3-week rounds over the span of 6 months, resulting in 9268 predictions (i.e., 4634 prediction pairs or sets). We focused on predicting financial prices as doing so is a hard prediction problem [62,63]. Our participants were mid-career financial professionals with years of financial experience. Our participants consented to their data being used in this study and we obtained prior IRB approval. One of our rounds of prediction happened to end the day of the Brexit vote, which means that we have prediction data during a particularly volatile market period [31] as described in Supplementary Section A.5 .

During each round, participants made a prediction of the same asset's closing price for the same final day of the round. We use the round's last day's closing market price as our measure of ground-truth. We carefully instrumented the social and non-social information that our participants were exposed to, and collected their predictions before and after exposure to this information. We also deployed one of our rounds during a high uncertainty period to understand if variance reduction strategies allow the crowd to be resistant to risk.

We did not opt for an A/B testing experimental design [64]—where we would have split participants and shown each group either the social information or the historical price time series—because we wanted participants to naturally choose whichever source of information to use to update their belief. This was an important experimental design choice as we wanted to understand, as close as possible to in-situ how people update their beliefs in the real-world where they are already exposed to both their peers' beliefs and to price history information, such as through financial news. Our design is in contrast to previous work where the experiments were deployed within a carefully controlled laboratory set-up as in prior work [25,37,40].

3.2. Data Collection

As shown in the screenshot of the user interface in Figure 1, we designed the data collection process as follows: every time a participant makes a prediction of an asset's future price through our platform, the following prediction set comprising B_{pre}, B_H, B_T and B_{post} is collected:

- A "pre-exposure" belief prediction B_{pre}, which is independent of *both* social information and price history. For example, a participant might show-up on the platform and predict that the closing price of the S&P 500 to be 2001 on 24 June 2016.
- The predictions B_H within the social information histogram shown to each participant after each initial prediction. Additionally, we display a 6-month time series of the asset's price B_T up to this point.
- The revised "post-exposure" prediction B_{post}. For example, after seeing the social histogram and asset price history, a participant might update their belief to 2201. Since the real price (the ground-truth V) ended up being 2037.41, this participant became more accurate after information exposure (they went from 2001 to 2201).
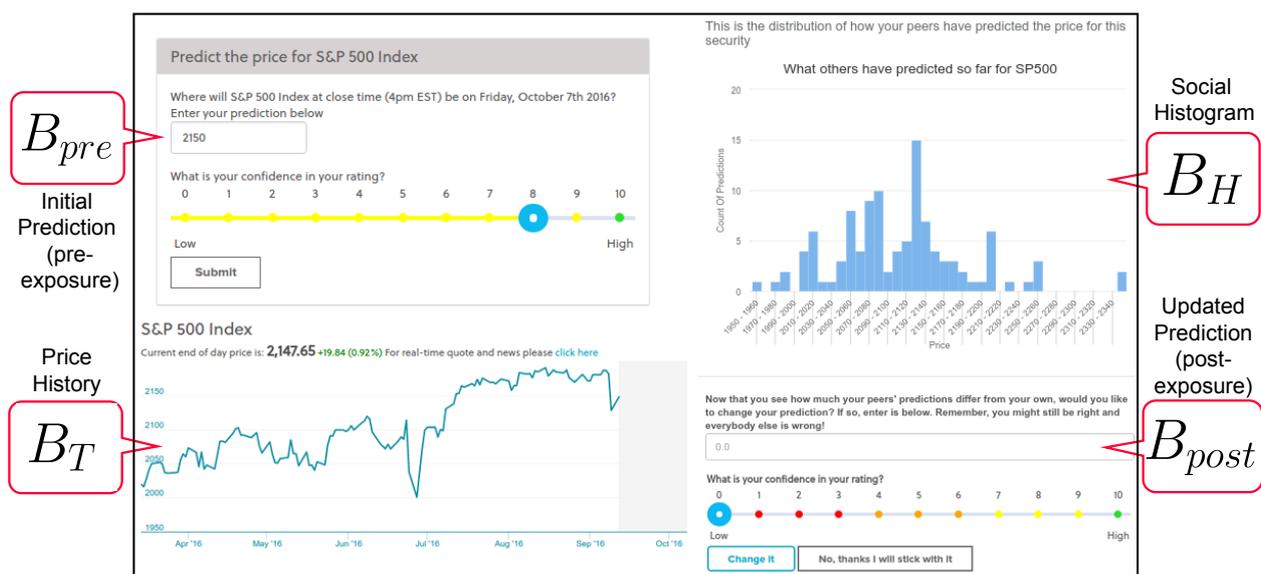

Figure 1. An annotated screenshot of how data were collected: the pre-exposure prediction B_{pre} is shown first, followed by the social histogram B_H and the price history B_T. Finally, the updated prediction B_{post} is collected. The ground-truth of the asset's final closing price will be V (not shown here, realized at the end of the round).

Overall, we ensure that the "pre-exposure" prediction is made before any social information and price history is shown. We present a unique histogram for every new prediction (as it is built using past predictions up to this point), as well as a unique price history time series (as it shows the 6-month price data up to the time of prediction). We require all participants to make a post-exposure prediction even if they decide to keep it at the pre-exposure level.

3.3. Modeling and Estimation

Using the data collected in the live experiments, we want to test our hypothesis that a Pareto frontier exists between risk and accuracy and that it is mediated by social learning. In this section, we describe all the modeling and estimation steps required to investigate our hypothesis:

- In Section 3.3.1, we describe how we model individual belief update: how a participant updates their prediction from a pre-exposure belief to a post-exposure prediction using a variety of models that are either Monte Carlo methods or simpler approximate methods inspired by Bayesian models of cognition [28,29]. This allows us to understand how participants update their belief after information exposure.
- In Section 3.3.2, using the models described earlier, we detail how to estimate the relative amount of social vs. non-social learning for each prediction to understand how much social vs. non-social data were factored into a prediction's belief update. We then introduce our methodology for selecting predictions based on the estimated amount

of social vs. non-social learning. This allows us to make aggregate predictions—at the platform level—based on a pre-specified amount of social learning.
- In Section 3.3.3, we detail how the accuracy and risk—at the platform level—of selected subsets are measured, and how they are used to investigate whether a Pareto trade-off exists between accuracy and risk and whether it is mediated by the relative amount of social vs. non-social learning.

3.3.1. Modeling Belief Updates

Using formalism inspired by Bayesian models of cognition [29], we can model the 4634 prediction sets collected over many rounds, at a high level, as a Bayesian update. To use this formalism, we need to select a prior distribution for each individual's belief before exposure to any information and a likelihood (evidence) distribution to model the data participants are exposed to. Additionally, a sampling or approximate method is required to use the prior and evidence to compute the posterior (updated belief after information exposure) distribution. Here, we describe the modeling assumptions and procedure at a high level, and detail more thoroughly our modeling assumptions and present our derivations in Supplementary Section A.3.

Fundamentally, we are interested in how participants predict an asset's future price (ground-truth) V based on the information we expose them to. The choice of the prior distribution is straightforward: $P_{prior}(V) \approx P(B_{pre})$, the distribution of belief of an individual before they are exposed to any information. We discuss in our model derivation (Supplementary Section A.3) how, when needed, we approximate the full distribution $P(B_{pre})$ since we obtain only one sample, B_{pre}, for each participant and cannot observe the full distribution $P(B_{pre})$.

After participants input their pre-exposure belief B_{pre}, there are two main likelihood (evidence) distributions participants employ: they are exposed to the assets' price history B_T, giving us $P_{likelihood}(V) \approx P(B_T)$, or analogously, the social histogram B_H, giving us $P_{likelihood}(V) \approx P(B_H)$. In the modeling stage here, we assume that participants used these two likelihood distributions separately to update their beliefs, but we relax this assumption in the estimation stage next where we estimate the relative amount of social vs. non-social learning for each prediction. We detail in Supplementary Section A.3 how likelihood distributions are built from the information that participants are exposed to. In Supplementary Section A.2, we formally detail how we transform the price history into a cognitively accurate 'rates histogram' using price momentum. As a summary, because it has been shown that people process time series as a distribution of changes as opposed to a distribution of the quantity itself [65–67], we convert the price history time series into a histogram of daily changes (slopes) in prices which is used for both the simple Gaussian models and the numerical models for price prediction.

Given the prior and likelihood, the *modeled* posterior prediction $P_{posterior}(V)$, can, therefore, be approximated as $P_{posterior}(V) \propto P(B_H) \cdot P(B_{pre})$ in the case of exposure to social information, and $P_{posterior}(V) \propto P(B_T) \cdot P(B_{pre})$ when participants are exposed to the past price history. We do not make any other assumptions in terms of what data to use to approximate the likelihood and prior distributions. Given these distributions, the question is then how to compute the posterior (updated) belief of an individual.

Although we focus on Bayesian models in this work, we include one popular model commonly used as a benchmark in the literature, the DeGroot model [68]. In this model, an individual updates their belief as the weighted average belief of their peers where weights can be, for example, trust values of the individual for their peers. Here we set the weights (trust values) equal for all peers, as we have no data to estimate these weights, and therefore assume a uniform prior.

Although the space of possible distributions and posterior computation approaches is very large, we focus here on using two simple, interpretable, and theoretically motivated approaches from prior work [28]. We either use Gaussian (normal) conjugate distributions to approximate priors and likelihoods due to strong evidence of their ubiquity as Bayesian

models of cognition [29], or use a full Monte Carlo numerical sampling approach to calculate the posterior from the actual distributions of prices that participants were exposed to. We leave to future work the exploration of richer distributions and approaches to modeling belief update as it is beyond the scope of this study.

3.3.2. Subsetting Predictions Based on Social Learning

Based on how participants update their belief, we would like to select subsets of predictions based on whether they were more likely updated using social or non-social information. This approach of using characteristics of how predictions are updated is standard in the Wisdom of the Crowd literature. For example, prior work has estimated resistance to social influence [27] and influenceability in revising judgments after seeing the opinion of others [69,70], and used them to improve collective performance. No prior work has investigated investigating if the modeling of belief update strategies could be leveraged for improved collective performance.

Using the previously modeled posteriors, we can *estimate* how much of each information source—social information and price history—each participant used to update their belief by comparing the residual errors of models using either only social information or only price history as likelihood. As will be introduced in the Section 4, although we explored many models of belief update, the simple conjugate Gaussian models model best how participants update their belief. This is in line with previous research showing that although simple, they are highly accurate models of mental estimation in a variety of domains [28].

Therefore, for the purposes of selecting subsets of prediction based on their relative amount of social vs. non-social learning, we choose to focus on the `GaussianSocial` and `GaussianPrice`. These models assume the likelihood (evidence) data distribution to be built, respectively, from the social information and price history participants are exposed to.

Our approach is illustrated in Figure 2: using the prediction of the models `GaussianSocial` and `GaussianPrice`, we calculate a residual ϵ_H for when updating belief using social information B_H and a residual ϵ_T when updating from the price history B_T, as $\epsilon_H = \frac{|\texttt{GaussianSocial} - B_{post}|}{B_{post}}$ and $\epsilon_T = \frac{|\texttt{GaussianPrice} - B_{post}|}{B_{post}}$ respectively. We define $\alpha = \epsilon_T - \epsilon_H$, and we use it to measure how likely a participant used each source of information to update their prediction. For example, for a prediction set $[B_{pre}, B_H, B_T, B_{post}]$ if $\alpha > 0$ (i.e., $\epsilon_T > \epsilon_H$), this means that this prediction set is better modeled using the social histogram of peer's belief B_H instead of the price history B_T.

Using α, which we re-scale to be in the interval $[-1, 1]$ for each round, we can select a subset S_{α_s} of the prediction sets such that the α of these prediction sets lie in the range $0 \leq \alpha < \alpha_s$ (or $\alpha_s < \alpha \leq 0$ when $\alpha_s < 0$). α_s is the one-sided boundary we will vary to measure how much more likely a participant updated their belief from the social information instead of the price history. For example, the higher α_s is, the more likely a prediction set is better modeled using the social histogram of peer's belief B_H instead of the price history B_T.

It is important to note that the residuals we use to select subsets are belief update model residuals (between the observed updated belief and the predicted modeled updated belief) which are uncorrelated with the crowd residual (between the crowd's aggregate prediction and the ground-truth).
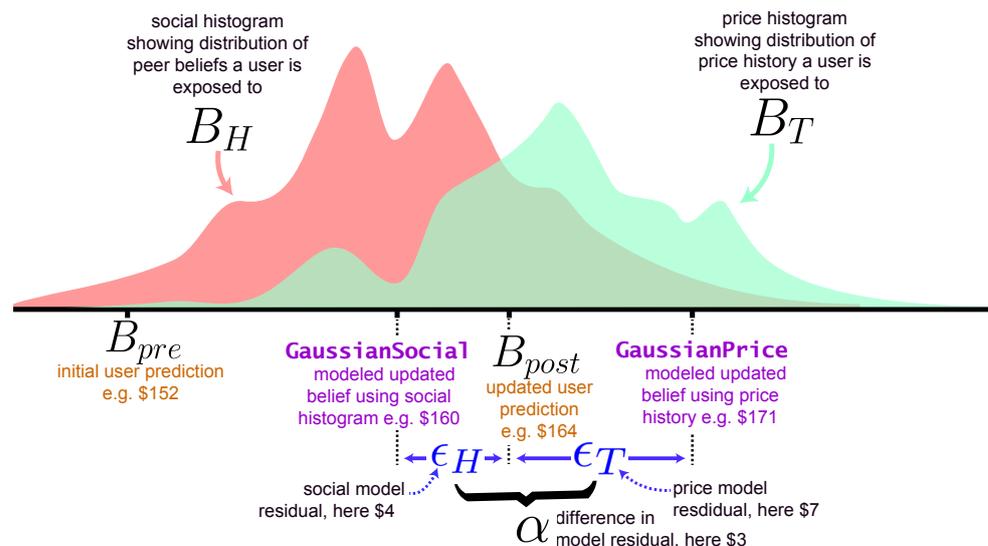
Figure 2. An example belief update: for each prediction set, a participant updates their belief from the pre-exposure prediction B_{pre} to the updated prediction B_{post} by either learning from the social histogram B_H and/or the price history B_T. ϵ_H is the residual between the *modeled* updated prediction `GaussianSocial` and the participant's updated prediction B_{post}; ϵ_T is the residual between `GaussianPrice` and B_{post}. α is the difference between ϵ_T and ϵ_H.

3.3.3. Evaluating Improvement of Subsets

Our hypothesis is that a Pareto frontier exists between risk and accuracy and that this trade-off is mediated by the relative amount of social vs. non-social learning.

To test this hypothesis, we investigate how the accuracy and variance of subsets S_{α_s} of predictions selected using α_s (a measure of the relative amount of social vs non-social learning) compares to the current standard Wisdom of the Crowd approach whereby all predictions are used.

From the perspective of platform designers who want to be able to select predictions based on required levels of accuracy or risk (e.g., to fit a certain portfolio of risk), it is important to measure improvement of subsets relative to the full collection of predictions. This is because, currently, platform designers only have access to one global measure of risk and accuracy—that of the whole set of predictions (when there is no subset filtering). To demonstrate that selecting subsets of predictions can lead to significant *improvements* in accuracy and risk, we therefore need to calculate these improvements.

We therefore define improvement $I^{S_{\alpha_s}}$ as the absolute difference between the error $e_{S_{\alpha_s}}$ when using a subset S_{α_s} compared to the error $e_{S_{all}}$ when using the full set of predictions S_{all}, the Wisdom of the Crowd, where S_{all} is defined as the full subset over all predictions using $-1 \leq \alpha \leq 1$.

The error $e_{i,S_{\alpha_s}}$ over all predictions $j \in S_{\alpha_s}$ for an estimated amount α_s of relative social vs. non-social information during experiment round i is defined as $\frac{|\sum_{j \in S_{\alpha_s}}[B_{post,j}] - V_i|}{V_i}$. To allow for estimation uncertainty over the improvement in accuracy and risk of subsets, we use 100 bootstraps with replacement. This procedure is formally described in Supplementary Section A.3.4.

We use an analogous approach to estimate the risk of the platform by calculating the standard deviation instead of the mean of the improvements over experiment rounds. This measures the risk for platform designers to estimate, over a basket of prediction rounds, what is the variance of improvements over this basket. This is the same as understanding the variance of error of a statistical prediction model (e.g., machine learning model) such that we can calibrate both the accuracy and variance of the model over a portfolio of predictions.

4. Results

Here we present our results. In Section 4.1, we detail our supporting result related to how different belief update models perform. Next, in Section 4.2, we present our main result about the trade-off between accuracy and risk in the Wisdom of the Crowd. Lastly, we present the supporting result regarding the effect of social learning during the high uncertainty period before the Brexit vote in Section 4.3.

4.1. Belief Update Models

Although the space of possible prior and likelihood distributions and posterior computation approaches is very large, we focus on using simple, interpretable, and theoretically motivated approaches from prior work [28]. We leave to future work the exploration of richer distributions and approaches to modeling belief update as it is beyond the scope of this study. We detail how model error and confidence intervals are evaluated in Supplementary Section A.3.3.

As can be seen in Figure 3, models that use social information as likelihood for modeling the belief update of participants (`GaussianSocial`,`GaussianSocialModes`, `Numerical Social`) outperform better than models that use the price history (`GaussianPrice`, `Numerical Price`). This suggests that our participants more likely use social information instead of the price history to update their belief, in line with previous work showing that participants often prefer using social information [71,72].
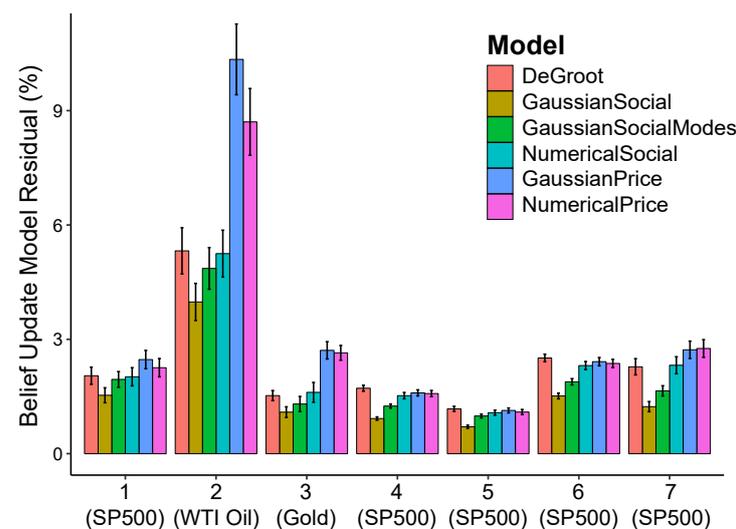

Figure 3. The y-axis shows the relative residual between *modeled* belief update and *actual* updated belief. Simple approximated models do better at modeling belief update than numerical models, and models using social histograms as likelihood perform better than models using the price history. Error bars represent 95% CI.

Specifically, `GaussianSocial`, our simple Gaussian model that assumes the data follows a single-mode Gaussian distribution, outperforms `GaussianSocialModes`, a model that identifies when the social histogram is non-unimodal (using the Hartigan's dip test of unimodality [73]) and uses the largest mode as the mean of the distribution. This suggests that participants assume the data they learn from to be unimodal even when it is non-unimodal, in line with prior work [74,75] showing that this might be due to the fact that using multi-modal data is cognitively costly.

Additionally, `GaussianSocial` outperforms the more precise numerical model `NumericalSocial` which makes no parametric assumption on the data distributions and uses a Monte Carlo procedure to estimate the posterior distribution. This suggests that participants employ simple heuristics when learning from their peers, in line with the attribute substitution heuristic of human decision-making [30]. However, when participants are learning from the price history, the dominance of simpler models is not as clear because

the performance of the simple `GaussianPrice` model is indistinguishable from that of the numerical model (`NumericalPrice`).

`GaussianSocial` also outperforms the popular `DeGroot` model commonly used as a benchmark in the literature [68], where an individual updates their belief as the weighted average belief of their peers. Here we set the weights (trust values) equal for all peers, as we have no data to estimate these weights, and therefore assume a uniform prior. It is interesting to note that `GaussianSocial` is equivalent to the `DeGroot` model when a participant's weight on their own prior belief is equal to the total of the weights of all other participants. This agrees with previous work showing that participants put a disproportionately larger weight on their own prior belief [76,77].

Overall, the superiority of `GaussianSocial` in predicting belief update suggests that participants use a heuristic, unimodal, and simple belief update procedure when updating their beliefs, and that they predominantly update their predictions using social information instead of price history. It is important to note that approximate (non-Monte Carlo) models such as `GaussianSocial` and `GaussianPrice` are parameter-less models and did not require any parameter fitting, making their success in modeling belief update quite interesting.

4.2. Accuracy-Risk Trade-Off

Here, we present our main result about the trade-off between accuracy and risk in the Wisdom of the Crowd. Using a Pareto curve, we compare the improvement in prediction accuracy and risk (variance) of each subset S_{α_s} as defined by α_s, a measure of the relative amount of social vs non-social learning.

As shown in Figure 4, we observe that with improvements in accuracy of subsets comes increased risk, mediated by the relative amount of social vs. non-social learning α_s, suggesting a trade-off between accuracy and risk. As formally described earlier in Section 3.3.3, improvement is a measure of the additional accuracy gained from a subset of predictions compared to when using all predictions by the crowd (the de-facto Wisdom of the Crowd) over all prediction rounds. Similarly, risk is a measure of the risk of this subset compared to when using all predictions over all rounds. From a system design perspective, we choose these measures of improvement and risk as they allow us to understand how choices over subsets of participants might affect performance, allowing us to calibrate the crowd as per the platform designer's risk preferences.
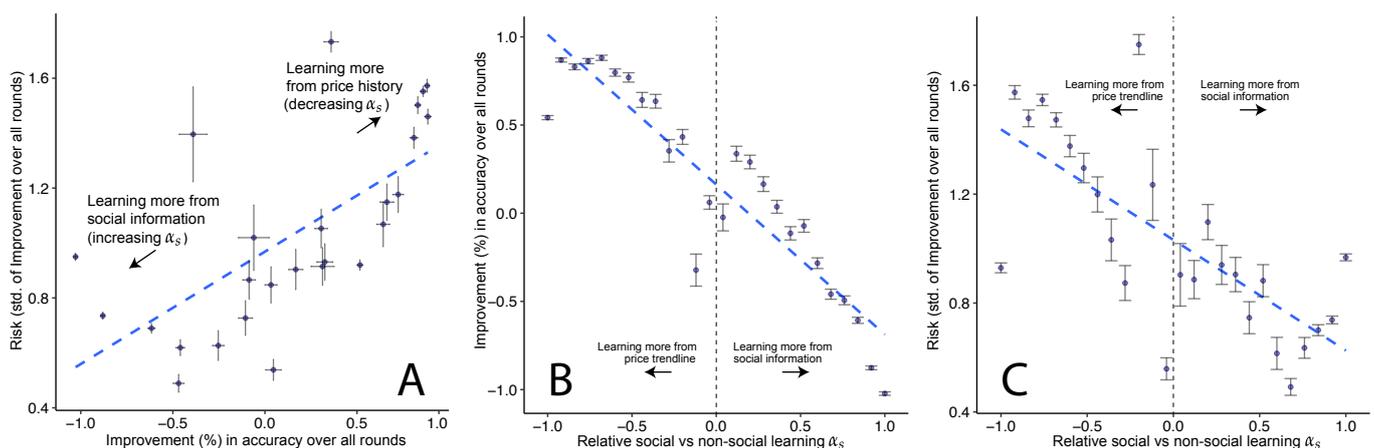

Figure 4. (**A**): In this Pareto curve, we plot the improvement of each subset vs. the risk (standard deviation) in improvement within this subset. We see a risk-return trade-off: predictions made with price history are more accurate, but with higher risk (standard deviation). Fitted line has R^2 of 0.49, and *p*-value < 0.001. Horizontal and vertical error bars represent 95% CI from 100 bootstraps. (**B**,**C**): Instead of plotting risk vs. improvement (as in (**A**)), here we plot the same values of improvement ((**B**), $R^2 = 0.82$, *p*-value < 0.001) or risk ((**C**), $R^2 = 0.50$, *p*-value < 0.001) against the relative amount of social vs. non-social learning, α_s, that generated these values of improvements or risk.

Additionally, since we observe that variance of improvement (risk) decreases with increased social leaning, our result replicates prior findings that exposure to social information decreases the variance of the crowd [37]). Please note that the decrease in risk from social learning is not because participants are simply converging towards the crowd's mean: as detailed in the previous Section 4.1, the social histogram participants are shown is quite often non-unimodal (tested using the Hartigan's dip test of unimodality [73]), which means that participants are intentionally collapsing multiple distribution modes in the observed data.

Such a Pareto trade-off between risk and accuracy is common in financial forecasting [15,16] and statistical prediction [8–11], but has not been typically observed in the literature on the Wisdom of Crowds. This has strong implications for the design of crowdsourced prediction platforms as described in the Discussion Section 5.1.

4.3. Performance under High Uncertainty

A supporting result of our work is from the investigation of the crowd's performance during a period of high uncertainty using the data from the prediction round that happened during the Brexit vote (see supplementary Section A.5 for details about this round).

Following the same procedure described in the Methods Section 3.3.3, we bin all α's from the prediction sets and investigate the improvements of subsets of predictions compared to the whole crowd. The main difference here is that unlike in all previous results where we took care not to use the last week of data to calculate collective accuracy so that prediction was not too easy, we do so here as the high uncertainty only happened in the last week (as shown in supplementary Figure S1). This last week of data that we use is a *disjoint subset* from the data we previously used.

As can be seen in Figure 5, as α_s decreases (i.e., we select predictions that were more likely updated using the price history instead of the social information, $\alpha_s < 0$), improvement in accuracy of subsets compared to the Wisdom of the Crowd (all predictions) decays to a great extent.
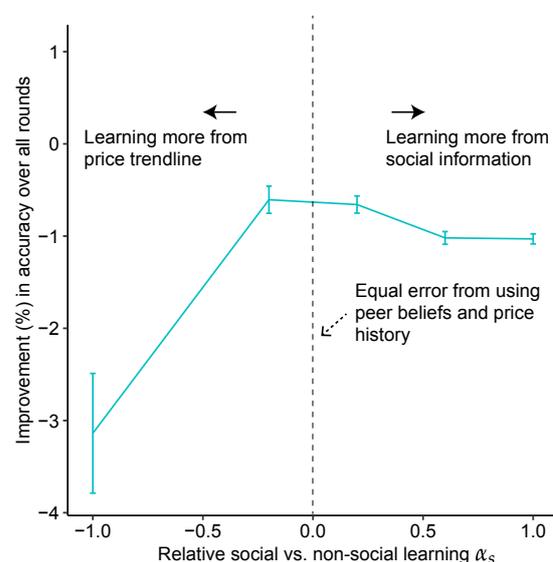

Figure 5. Improvement when selecting predictions based on how much more they were likely made using social information ($\alpha_s > 0$) vs. price history ($\alpha_s < 0$). 95% Confidence intervals obtained through 100 bootstraps.

Conversely, as subsets of predictions updated using the social histogram ($\alpha_s > 0$) are selected, the improvement in their accuracy is stable.

Given that such high market uncertainty only occurred during one round, we do not have enough data to produce a Pareto curve over multiple rounds. Additionally, note that although a smaller number of predictions were made during the last week before Brexit

(52 prediction sets compared to 284 during the open period of prediction used earlier), we have sufficient data to afford statistically significant results as shown by the 95% confidence intervals of our findings.

This supporting result suggests that during periods of high uncertainty, social learning leads to higher accuracy in contrast to the result in the previous section where the asset prices were more predictable. This result has implications for platform designers such as the potential of leveraging social learning as a valuable tool that minimizes catastrophic performance during high uncertainty prediction regimes.

5. Discussion

Our main result (the trade-off seen in Figure 4) supports our hypothesis that a Pareto frontier exists between risk and accuracy—similarly to what has been observed in statistical modeling [8–10] and financial [14–16] forecasting systems. This trade-off is mediated by the relative amount of social vs. non-social learning. Additionally, as supporting results, we observe that simple approximate models outperform more complicated Monte Carlo approaches in modeling the belief update process of participants. This suggests that participants use several heuristics, and that during periods of high uncertainty, social learning leads to higher accuracy.

Here, we discuss the implications of our results for platform designers in Section 5.1, describe the contributions of our work to the literature on heuristics in information processing and decision-making in Section 5.2. We end with a description the limitations of this work in Section 5.3.

5.1. Collective Intelligence System Design Implications

If we are to deploy crowd-sourced financial prediction and speculation systems at scale, it will be important to fully characterize the performance of these systems. This is especially given the growing importance of decentralized financial prediction and speculation including very recent events during which retail investors self-organized using social media and drove up asset and derivative prices [3,4]. However, crowd-sourced prediction systems and literature so far focus on measuring and optimizing for the accuracy of the predictions with little regard to the risk of these predictions even though measuring both accuracy and risk is standard in machine learning [8–10] and financial [14–16] forecasting applications. More generally, proper modeling and estimation of risk will support more sophisticated and versatile applications of crowd-sourced predictions such as hedging risks over portfolios of prediction tasks.

Additionally, beyond the passive monitoring and reporting of risk, a practical question for designers is how to *tune* the platform to reach a desired value of risk and accuracy. Our result that social learning can mediate the accuracy-risk trade-off provides a practical means to attain performance along this frontier. Specifically, our results suggest that social learning within a crowd-sourcing platform could be more purposefully leveraged to fit the task at hand. For example, platform designers could incentivize social learning between participants to have lower risk. This might be especially needed during highly uncertain times, as our results from the Brexit prediction (Figure 5) prediction showed. Past work has already showed that crowd-sourcing platforms can be incentivized to be more social [43,44].

Beyond platform design considerations, our results also add to the rich study of social learning and its impact on collective intelligence within the Wisdom of the Crowd domain [25,27,37,40,41] by adding the novel perspective that risk is an important dimension of the behavior of crowds to be measured.

More generally, our work brings together two disjoint studies by showing that it is possible to improve collective intelligence by modeling individual belief update. Our results therefore suggest a connection between the field of collective intelligence [78] (of which the Wisdom of the Crowd is one domain) and the field of computational cognitive science [79] (of which Bayesian models of cognition is an area). Until now, the latter literature has mostly focused on individual models of belief update such as through

computational models of how people perform sampling [80], what their priors are [81], and how they perform inference [82], sometimes in social situations [83]. Yet, there is little work that looks at the impact of individual belief update on collective performance. On the other hand, there is limited collective intelligence literature regarding leveraging the modeling of individual belief update to improve group performance and past work has instead been focused on using personal characteristics such as resistance to social learning [27].

5.2. Information Processing and Decision-Making Heuristics

Our results also have implications for the literature on decision heuristics and biases [75,84]. Through the modeling of belief update, we observe that our subjects exhibit the attribute substitution heuristic of human decision-making [30]. This information processing heuristic describes when people attempt to solve a complicated problem by approximating it with a simpler, less accurate model. We observe this heuristic as our participants' updated beliefs are better modeled by the `GaussianSocial` model (which assumes the data to be unimodal) than by the multi-modal belief update model `GaussianSocialModes`. This indicates that our participants assume the data to be unimodal even when it is not, in line with previous studies that have shown that people wrongly assume data to be unimodal [74,85,86]. This is hypothesized to be because updating belief using multi-modal data is cognitively costly [87]. Additional evidence of this substitution heuristic is from the fact that simpler, approximate models better predict the updated beliefs of participants than the more complicated Monte Carlo numerical models.

Another decision heuristic that we observe is that participants prefer to use social information rather than the underlying price history of an asset to update their belief as models which use social information (`GaussianSocial`,`GaussianSocialModes`, and `NumericalSocial`) outperform models that use price history (`GaussianPrice` and `NumericalPrice`) as shown in Figure 3. This is surprising given that our participants were mid-career finance professionals with strong financial experience who should know that price information is generally better to predict future prices [88,89]). However, such behavior was observed in prior work where even experts performing a familiar task demonstrate sub-optimal decision heuristics [90,91], and often over-rely on social information [71,72].

Generally, such information processing and decision-making heuristics have been seen as irrational and sub-optimal. Our results suggest that within the full specification of both accuracy and risk, perhaps participants are preferentially aiming for lower risk instead of higher accuracy. This preference for social information especially pays off during the high uncertainty period before the Brexit vote. Our results support growing evidence that heuristics and biases are not merely *defects* of human decision-making, but that perhaps they optimize for richer objectives or are optimized for more time- or data-constrained decision-making [92–98]. For example, when individual decision-making is viewed within the lens of more realistic requirements such as limited time [99,100] or attention [101], heuristics and biases have been shown to act as helpful priors that facilitate fast and risk-averse decision-making [102,103].

5.3. Limitations and Future Work

We made several simplifying assumptions in this work that open up rich avenues for future work. First, we used simple, interpretable, and theoretically motivated belief update modeling approaches from prior work [28] and leave to future work the exploration of richer models, distributions and posterior computations to investigate belief update. One important set of models to investigate is the use of log-normal distributions for the likelihood instead of the normal distributions used in this work due to the established tendency of people to guess quantities log-normally [37,104,105]. Similarly, people have been shown to incorporate information asymmetrically based on where their predictions lie in relation to the information they are exposed to [106]. Overall, although we used Gaussian models here, an interesting direction of future work would be to build on the rich existing literature on how people incorporate information [84,107,108]. We also restricted

each round to have a static population of participants whose predictions were shared using a specific visualization. An interesting direction for future work would be to embed participants in social networks given the importance and popularity of recent work on the effect of communication topologies [25,41,42,109] on group performance. Similarly, it would be interesting to investigate if different avenues for communication (e.g., discussions on forums [110]) exhibit a similar accuracy-risk trade-off.

Although this work demonstrates that our simple estimation technique can be used to tune crowd-predictions for desired levels of accuracy and risk, there are potential causal issues that could be improved in our experimental design and data analysis. One such issue is that there are two experimental and two analysis factors being investigated simultaneously here. These are the two different treatments in the form of sources of information (peer beliefs for the social histogram and price trajectory from the past price history) and the two different approaches through which each of these sources of information are being processed (simple binning of peer beliefs into a histogram, and transformation of the price history into a 'rates histogram'). It can be argued that these two experimental treatments and two approaches constitute four possible approaches of how to deploy and analyze an experiment, and we have only compared two of these four approaches. From a scholarly perspective, we believe that our paper still makes a contribution because the goal of this work was to show that a trade-off exists and is mediated by social learning. We achieve this goal even though we only compare two approaches. Another causal concern is that the two experimental treatments might interact in non-trivial ways. For example, when visualized as a causal graph, there might be causally confounding paths between the treatments.

Several research designs and estimations techniques exist to remedy these causal limitations. One approach would be to use an A/B test [64] framework although it would require exposing people to different information separately. Doing so would be against our goal to investigate how people update their belief in real-life situations where users are exposed to both social information and price history. However, experiments where different types of information are shown separately could still be used to understand the effect of different information exposures on accuracy and risk, and used in deployment. Similarly different amounts of information exposure could be attempted using a multi-factorial A/B test [111,112]. We leave the exploration of these more sophisticated designs to future work. Other de-confounding approaches could involve assuming a causal graph [113] that is believed to capture how people update information and to use causal tools such as d-separation to estimate the effect of different information exposure. Another approach would be to use a potential outcomes framework [114] to estimate these treatments. These are promising directions of research which could be investigated using our data that we leave to future work. From a platform design perspective, even though these confounding issues remain, our estimation technique could be readily applied to crowd-sourced systems where price histories and peer beliefs are being shown.

Author Contributions: Conceptualization, D.A.; methodology, D.A., Y.L., S.K.C., P.M.K.; validation, D.A.; formal analysis, D.A., Y.L., S.K.C.; investigation, D.A., Y.L., S.K.C.; resources, A.P.; data curation, D.A., Y.L., S.C; writing—original draft preparation, all authors; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the MIT COUHES IRB and approved as Exempt Protocol 1602374158.

Informed Consent Statement: Study participants consented to their data being used in this study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Diallo, N.; Shi, W.; Xu, L.; Gao, Z.; Chen, L.; Lu, Y.; Shah, N.; Carranco, L.; Le, T.C.; Surez, A.B.; et al. eGov-DAO: A better government using blockchain based decentralized autonomous organization. In Proceedings of the 2018 International Conference on eDemocracy & eGovernment (ICEDEG), Ambato, Ecuador, 4–6 April 2018; pp. 166–171.
2. Lang, M.; Bharadwaj, N.; Di Benedetto, C.A. How crowdsourcing improves prediction of market-oriented outcomes. *J. Bus. Res.* **2016**, *69*, 4168–4176. [CrossRef]
3. Lawrence, K. Memes, Reddit, and Robinhood: Analyzing the GameStop Saga. 2021. Available online: http://sk.sagepub.com/cases/memes-reddit-and-robinhood-analyzing-the-gamestop-saga (accessed on 10 May 2021)
4. Hu, D.; Jones, C.M.; Zhang, V.; Zhang, X. The Rise of Reddit: How Social Media Affects Retail Investors and Short-Sellers' Roles in Price Discovery. 2021. Available online : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3807655 (accessed on 10 May 2021).
5. Lazer, D.; Friedman, A. The network structure of exploration and exploitation. *Adm. Sci. Q.* **2007**, *52*, 667–694. [CrossRef]
6. Olorunda, O.; Engelbrecht, A.P. Measuring exploration/exploitation in particle swarms using swarm diversity. In Proceedings of the 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008; pp. 1128–1134.
7. Kennedy, J.; Mendes, R. Population structure and particle swarm performance. In Proceedings of the 2002 Congress on Evolutionary Computation, CEC'02 (Cat. No. 02TH8600), Honolulu, HI, USA, 12–17 May 2002; Volume 2, pp. 1671–1676.
8. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.
9. Domingos, P. A unified bias-variance decomposition. In Proceedings of 17th International Conference on Machine Learning, Stanford, CA, USA, 29 June–2 July 2000 ; pp. 231–238.
10. Geman, S.; Bienenstock, E.; Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **1992**, *4*, 1–58. [CrossRef]
11. Gagliardi, F. Instance-based classifiers applied to medical databases: diagnosis and knowledge extraction. *Artif. Intell. Med.* **2011**, *52*, 123–139. [CrossRef] [PubMed]
12. Markowitz, H. Portfolio selection. *J. Financ.* **1952**, *7*, 77–91.
13. Gammerman, A.; Vovk, V. Hedging predictions in machine learning. *Comput. J.* **2007**, *50*, 151–163. [CrossRef]
14. Joyce, J.M.; Vogel, R.C. The uncertainty in risk: Is variance unambiguous? *J. Financ.* **1970**, *25*, 127–134. [CrossRef]
15. Modigliani, F.; Leah, M. Risk-adjusted performance. *J. Portf. Manag.* **1997**, *23*, 45. [CrossRef]
16. Ghysels, E.; Santa-Clara, P.; Valkanov, R. There is a risk-return trade-off after all. *J. Financ. Econ.* **2005**, *76*, 509–548. [CrossRef]
17. Chavez-Demoulin, V.; Embrechts, P.; Nešlehová, J. Quantitative models for operational risk: extremes, dependence and aggregation. *J. Bank. Financ.* **2006**, *30*, 2635–2658. [CrossRef]
18. Asmussen, S.; Kroese, D.P. Improved algorithms for rare event simulation with heavy tails. *Adv. Appl. Probab.* **2006**, *38*, 545–558. [CrossRef]
19. Shevchenko, P.V.; Wuthrich, M.V. The structural modelling of operational risk via Bayesian inference: Combining loss data with expert opinions. *J. Oper. Risk* **2006**, *1*, 3–26. [CrossRef]
20. Chapelle, A.; Crama, Y.; Hübner, G.; Peters, J.P. Practical methods for measuring and managing operational risk in the financial sector: A clinical study. *J. Bank. Financ.* **2008**, *32*, 1049–1061. [CrossRef]
21. Cruz, M.G. *Modeling, Measuring and Hedging Operational Risk*; Wiley: New York, NY, USA, 2002; Volume 346.
22. Galton, F. Vox populi (The wisdom of crowds). *Nature* **1907**, *75*, 450–451. [CrossRef]
23. Golub, B.; Jackson, M.O. Naive learning in social networks and the wisdom of crowds. *Am. Econ. J. Microecon.* **2010**, *2*, 112–149. [CrossRef]
24. Nofer, M.; Hinz, O. Are crowds on the internet wiser than experts? The case of a stock prediction community. *J. Bus. Econ.* **2014**, *84*, 303–338. [CrossRef]
25. Becker, J.; Brackbill, D.; Centola, D. Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 201615978; [CrossRef] [PubMed]
26. Turner, B.M.; Steyvers, M.; Merkle, E.C.; Budescu, D.V.; Wallsten, T.S. Forecast aggregation via recalibration. *Mach. Learn.* **2014**, *95*, 261–289. [CrossRef]
27. Madirolas, G.; de Polavieja, G.G. Improving collective estimations using resistance to social influence. *PLoS Comput. Biol.* **2015**, *11*, e1004594. [CrossRef]
28. Griffiths, T.L.; Tenenbaum, J.B. Optimal predictions in everyday cognition. *Psychol. Sci.* **2006**, *17*, 767–773. [CrossRef] [PubMed]
29. Griffiths, T.L.; Kemp, C.; Tenenbaum, J.B. Bayesian models of cognition. In *The Cambridge Handbook of Computational Psychology*; Sun, R., Ed.; Cambridge University Press: Cambridge, UK, 2008; pp. 1–49.
30. Kahneman, D.; Frederick, S. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. Available online: https://www.cambridge.org/core/books/heuristics-and-biases/representativeness-revisited-attribute-substitution-in-intuitive-judgment/AAB5D933A3F944CFB5CB02265D376C8F (accessed on 10 May 2021).

31. Oehler, A.; Horn, M.; Wendt, S. Brexit: Short-term stock price effects and the impact of firm-level internationalization. *Financ. Res. Lett.* **2017**, *22*, 175–181. [CrossRef]

32. Kennedy, J. Swarm intelligence. In *Handbook of Nature-Inspired and Innovative Computing*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 187–219.

33. Eberhart, R.C.; Shi, Y.; Kennedy, J. *Swarm Intelligence*; Elsevier: Amsterdam, The Netherlands, 2001.

34. Bonabeau, E.; Marco, D.D.R.D.F.; Dorigo, M.; Théraulaz, G.; Theraulaz, G. *Swarm Intelligence: From Natural to Artificial Systems*; Number 1; Oxford University Press: Oxford, UK, 1999.

35. Woolley, A.W.; Chabris, C.F.; Pentland, A.; Hashmi, N.; Malone, T.W. Evidence for a collective intelligence factor in the performance of human groups. *Science* **2010**, *330*, 686–688. [CrossRef]

36. Malone, T.W.; Laubacher, R.; Dellarocas, C. The collective intelligence genome. *MIT Sloan Manag. Rev.* **2010**, *51*, 21. [CrossRef]

37. Lorenz, J.; Rauhut, H.; Schweitzer, F.; Helbing, D. How social influence can undermine the wisdom of crowd effect. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9020–9025; [CrossRef]

38. Muchnik, L.; Aral, S.; Taylor, S.J. Social influence bias: A randomized experiment. *Science* **2013**, *341*, 647–651. [CrossRef]

39. Salganik, M.J.; Dodds, P.S.; Watts, D.J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **2006**, *311*, 854–856. [CrossRef] [PubMed]

40. Moussaïd, M.; Kämmer, J.E.; Analytis, P.P.; Neth, H. Social influence and the collective dynamics of opinion formation. *PLoS ONE* **2013**, *8*, e78433. [CrossRef] [PubMed]

41. Almaatouq, A.; Noriega-Campero, A.; Alotaibi, A.; Krafft, P.; Moussaid, M.; Pentland, A. Adaptive social networks promote the wisdom of crowds. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 11379–11386. [CrossRef]

42. Adjodah, D.; Calacci, D.; Dubey, A.; Goyal, A.; Krafft, P.; Moro, E.; Pentland, A. Leveraging Communication Topologies Between Learning Agents in Deep Reinforcement Learning. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, Auckland, New Zealand, 9–13 May 2020.

43. Lim, S.L.; Quercia, D.; Finkelstein, A. StakeSource: harnessing the power of crowdsourcing and social networks in stakeholder analysis. In Proceedings of the 2010 ACM/IEEE 32nd International Conference on Software Engineering, Cape Town, South Africa, 2–8 May 2010; Volume 2, pp. 239–242.

44. Chen, P.Y.; Cheng, S.M.; Ting, P.S.; Lien, C.W.; Chu, F.J. When crowdsourcing meets mobile sensing: A social network perspective. *IEEE Commun. Mag.* **2015**, *53*, 157–163. [CrossRef]

45. Lerman, K.; Ghosh, R. Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM), Washington, DC, USA, 23–26 May 2010.

46. Lerman, K.; Hogg, T. Using a model of social dynamics to predict popularity of news. In Proceedings of the 19th International Conference on World Wide Web (WWW), Raleigh, NC, USA, 26–30 April 2010; pp. 621–630.

47. Stoddard, G. Popularity dynamics and intrinsic quality in reddit and hacker news. In Proceedings of the Ninth International AAAI Conference on Web and Social Media (ICWSM), Oxford, UK, 26–29 May 2015.

48. Celis, L.E.; Krafft, P.M.; Kobe, N. Sequential voting promotes collective discovery in social recommendation systems. In Proceedings of the Tenth International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016.

49. Karger, D.R.; Oh, S.; Shah, D. Budget-optimal task allocation for reliable crowdsourcing systems. *Oper. Res.* **2014**, *62*, 1–24. [CrossRef]

50. Holt, C.A.; Laury, S.K. Risk aversion and incentive effects. *Am. Econ. Rev.* **2002**, *92*, 1644–1655. [CrossRef]

51. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making: Part I*; World Scientific: Singapore, 2013; pp. 99–127.

52. Hintze, A.; Olson, R.S.; Adami, C.; Hertwig, R. Risk sensitivity as an evolutionary adaptation. *Sci. Rep.* **2015**, *5*, 8242. [CrossRef]

53. Zhang, R.; Brennan, T.J.; Lo, A.W. The origin of risk aversion. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17777–17782. [CrossRef] [PubMed]

54. Binswanger, H.P.; Sillers, D.A. Risk aversion and credit constraints in farmers' decision-making: A reinterpretation. *J. Dev. Stud.* **1983**, *20*, 5–21. [CrossRef]

55. Armstrong, J.S.; Green, K.C.; Graefe, A. Golden rule of forecasting: Be conservative. *J. Bus. Res.* **2015**, *68*, 1717–1731. [CrossRef]

56. Passino, K.M.; Seeley, T.D. Modeling and analysis of nest-site selection by honeybee swarms: the speed and accuracy trade-off. *Behav. Ecol. Sociobiol.* **2006**, *59*, 427–442. [CrossRef]

57. Valentini, G.; Hamann, H.; Dorigo, M. Efficient decision-making in a self-organizing robot swarm: On the speed versus accuracy trade-off. In Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey, 4–8 May 2015; pp. 1305–1314.

58. Krause, J.; Ruxton, G.D.; Krause, S. Swarm intelligence in animals and humans. *Trends Ecol. Evol.* **2010**, *25*, 28–34. [CrossRef]

59. Wolf, M.; Kurvers, R.H.; Ward, A.J.; Krause, S.; Krause, J. Accurate decisions in an uncertain world: Collective cognition increases true positives while decreasing false positives. *Proc. R. Soc. B Biol. Sci.* **2013**, *280*, 20122777. [CrossRef]

60. Ward, A.J.; Herbert-Read, J.E.; Sumpter, D.J.; Krause, J. Fast and accurate decisions through collective vigilance in fish shoals. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2312–2315. [CrossRef]

61. Dumas, B.; Fleming, J.; Whaley, R.E. Implied volatility functions: Empirical tests. *J. Financ.* **1998**, *53*, 2059–2106. [CrossRef]

62. Campbell, J.Y.; Shiller, R.J. Stock prices, earnings, and expected dividends. *J. Financ.* **1988**, *43*, 661–676. [CrossRef]

63. Fama, E.F. Random walks in stock market prices. *Financ. Anal. J.* **1995**, *51*, 75–80. [CrossRef]
64. Dixon, E.; Enos, E.; Brodmerkle, S. A/b Testing of a Webpage. U.S. Patent 7,975,000, 2013. Available online: https://patents.google.com/patent/US20060162071A1/en (accessed on 10 May 2021)
65. Maniadakis, M.; Trahanias, P. Time models and cognitive processes: A review. *Front. Neurorobotics* **2014**, *8*, 7. [CrossRef] [PubMed]
66. Park, C.H.; Irwin, S.H. What do we know about the profitability of technical analysis? *J. Econ. Surv.* **2007**, *21*, 786–826. [CrossRef]
67. Neftci, S.N. Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of "technical analysis". *J. Bus.* **1991**, *64*, 549–571. [CrossRef]
68. DeGroot, M.H. Reaching a consensus. *J. Am. Stat. Assoc.* **1974**, *69*, 118–121. [CrossRef]
69. Kerckhove, C.V.; Martin, S.; Gend, P.; Rentfrow, P.J.; Hendrickx, J.M.; Blondel, V.D. Modelling influence and opinion evolution in online collective behaviour. *PLoS ONE* **2016**, *11*, e0157685. [CrossRef]
70. Soll, J.B.; Larrick, R.P. Strategies for revising judgment: How (and how well) people use others' opinions. *J. Exp. Psychol. Learn. Mem. Cogn.* **2009**, *35*, 780. [CrossRef]
71. Foster, F.D.; Viswanathan, S. Strategic trading when agents forecast the forecasts of others. *J. Financ.* **1996**, *51*, 1437–1478. [CrossRef]
72. Posada, M.; Hernandez, C.; Lopez-Paredes, A. Learning in continuous double auction market. In *Artificial Economics*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 41–51.
73. Hartigan, J.A.; Hartigan, P.M. The dip test of unimodality. *Ann. Stat.* **1985**, *13*, 70–84. [CrossRef]
74. Donnelly, N.; Cave, K.; Welland, M.; Menneer, T. Breast screening, chicken sexing and the search for oil: Challenges for visual cognition. *Geol. Soc. Lond. Spec. Publ.* **2006**, *254*, 43–55. [CrossRef]
75. Nisbett, R.E.; Ross, L. Human Inference: Strategies and Shortcomings of Social Judgment. 1980. Available online: https://philpapers.org/rec/nishis (accessed on 10 May 2021).
76. Dave, C.; Wolfe, K.W. On confirmation bias and deviations from Bayesian updating. *Retrieved* **2003**, *24*, 2011.
77. Nickerson, R.S. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **1998**, *2*, 175–220. [CrossRef]
78. Mataric, M.J. Designing emergent behaviors: From local interactions to collective intelligence. In Proceedings of the Second International Conference on Simulation of Adaptive Behavior, Honolulu, HI, USA, 13 April 1993; pp. 432–441.
79. Tenenbaum, J.B.; Kemp, C.; Griffiths, T.L.; Goodman, N.D. How to grow a mind: Statistics, structure, and abstraction. *Science* **2011**, *331*, 1279–1285. [CrossRef] [PubMed]
80. Vul, E.; Pashler, H. Measuring the crowd within probabilistic representations within individuals. *Psychol. Sci.* **2008**, *19*, 645–647. [CrossRef]
81. Lewandowsky, S.; Griffiths, T.L.; Kalish, M.L. The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. *Cogn. Sci.* **2009**, *33*, 969–998. [CrossRef]
82. Tenenbaum, J.B.; Griffiths, T.L.; Kemp, C. Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* **2006**, *10*, 309–318. [CrossRef] [PubMed]
83. Baker, C.L.; Saxe, R.; Tenenbaum, J.B. Action understanding as inverse planning. *Cognition* **2009**, *113*, 329–349. [CrossRef]
84. Tversky, A.; Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science* **1974**, *185*, 1124–1131. [CrossRef]
85. Nisbett, R.E.; Kunda, Z. Perception of social distributions. *J. Personal. Soc. Psychol.* **1985**, *48*, 297. [CrossRef]
86. Lindskog, M. Is the Intuitive Statistician Eager or Lazy?: Exploring the Cognitive Processes of Intuitive Statistical Judgments. Ph.D. Thesis, Acta Universitatis Upsaliensis, 2013. Available online: https://www.diva-portal.org/smash/record.jsf?pid=diva2 (accessed on 10 May 2021)
87. Hoffman, A.B.; Rehder, B. The costs of supervised classification: The effect of learning task on conceptual flexibility. *J. Exp. Psychol. Gen.* **2010**, *139*, 319. [CrossRef]
88. Malkiel, B.G.; Fama, E.F. Efficient capital markets: A review of theory and empirical work. *J. Financ.* **1970**, *25*, 383–417. [CrossRef]
89. Fama, E.F. The behavior of stock-market prices. *J. Bus.* **1965**, *38*, 34–105. [CrossRef]
90. Shanteau, J. Psychological characteristics and strategies of expert decision makers. *Acta Psychol.* **1988**, *68*, 203–215. [CrossRef]
91. Koehler, D.J.; Brenner, L.; Griffin, D. The calibration of expert judgment: Heuristics and biases beyond the laboratory. *Heuristics Biases Psychol. Intuitive Judgm.* **2002**, 686–715. Available online: https://psycnet.apa.org/record/2003-02858-039 (accessed on 10 May 2021).
92. Lakshminarayanan, V.R.; Chen, M.K.; Santos, L.R. The evolution of decision-making under risk: Framing effects in monkey risk preferences. *J. Exp. Soc. Psychol.* **2011**, *47*, 689–693. [CrossRef]
93. Mallpress, D.E.; Fawcett, T.W.; Houston, A.I.; McNamara, J.M. Risk attitudes in a changing environment: An evolutionary model of the fourfold pattern of risk preferences. *Psychol. Rev.* **2015**, *122*, 364. [CrossRef] [PubMed]
94. Kenrick, D.T.; Griskevicius, V. *The Rational Animal: How Evolution Made Us Smarter than We Think*; Basic Books (AZ): 2013. Available online: https://psycnet.apa.org/record/2013-31943-000 (accessed on 10 May 2021)
95. Josef, A.K.; Richter, D.; Samanez-Larkin, G.R.; Wagner, G.G.; Hertwig, R.; Mata, R. Stability and change in risk-taking propensity across the adult life span. *J. Personal. Soc. Psychol.* **2016**, *111*, 430. [CrossRef] [PubMed]
96. Cronqvist, H.; Siegel, S. The genetics of investment biases. *J. Financ. Econ.* **2014**, *113*, 215–234. [CrossRef]
97. Santos, L.R.; Rosati, A.G. The evolutionary roots of human decision making. *Annu. Rev. Psychol.* **2015**, *66*, 321–347. [CrossRef] [PubMed]

98. Mishra, S. Decision-making under risk: Integrating perspectives from biology, economics, and psychology. *Personal. Soc. Psychol. Rev.* **2014**, *18*, 280–307. [CrossRef]

99. Azuma, R.; Daily, M.; Furmanski, C. A review of time critical decision making models and human cognitive processes. In Proceedings of the 2006 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2006; p. 9.

100. Cohen, I. Improving time-critical decision making in life-threatening situations: Observations and insights. *Decis. Anal.* **2008**, *5*, 100–110. [CrossRef]

101. Van Knippenberg, D.; Dahlander, L.; Haas, M.R.; George, G. Information, Attention, and Decision Making. 2015. Available online: https://psycnet.apa.org/record/2015-33332-001 (accessed on 10 May 2021)

102. Lubell, M.; Scholz, J.T. Cooperation, reciprocity, and the collective-action heuristic. *Am. J. Political Sci.* **2001**, *45*, 160–178. [CrossRef]

103. Rand, D.G.; Brescoll, V.L.; Everett, J.A.; Capraro, V.; Barcelo, H. Social heuristics and social roles: Intuition favors altruism for women but not for men. *J. Exp. Psychol. Gen.* **2016**, *145*, 389. [CrossRef]

104. Limpert, E.; Stahel, W.A.; Abbt, M. Log-normal distributions across the sciences: keys and clues: on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—Normal or log-normal: that is the question. *BioScience* **2001**, *51*, 341–352.

105. Dehaene, S.; Izard, V.; Spelke, E.; Pica, P. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* **2008**, *320*, 1217–1220. [CrossRef] [PubMed]

106. Kao, A.B.; Berdahl, A.M.; Hartnett, A.T.; Lutz, M.J.; Bak-Coleman, J.B.; Ioannou, C.C.; Giam, X.; Couzin, I.D. Counteracting estimation bias and social influence to improve the wisdom of crowds. *J. R. Soc. Interface* **2018**, *15*, 20180130. [CrossRef]

107. Payne, J.W.; Payne, J.W.; Bettman, J.R.; Johnson, E.J. *The Adaptive Decision Maker*; Cambridge University Press: Cambridge, UK, 1993.

108. Schneider, W.; Shiffrin, R.M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* **1977**, *84*, 1. [CrossRef]

109. Barkoczi, D.; Galesic, M. Social learning strategies modify the effect of network structure on group performance. *Nat. Commun.* **2016**, *7*, 1–8. [CrossRef] [PubMed]

110. Krafft, P.M.; Della Penna, N.; Pentland, A.S. An experimental study of cryptocurrency market dynamics. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.

111. Plackett, R.L.; Burman, J.P. The design of optimum multifactorial experiments. *Biometrika* **1946**, *33*, 305–325. [CrossRef]

112. Montgomery, D.C. *Design and Analysis of Experiments*; John Wiley & Sons: Hoboken, NJ, USA, 2017.

113. Pearl, J. Causal diagrams for empirical research. *Biometrika* **1995**, *82*, 669–688. [CrossRef]

114. Rubin, D.B. Causal inference using potential outcomes: Design, modeling, decisions. *J. Am. Stat. Assoc.* **2005**, *100*, 322–331. [CrossRef]

115. Alquist, R.; Kilian, L. What do we learn from the price of crude oil futures? *J. Appl. Econom.* **2010**, *25*, 539–573. [CrossRef]

116. French, K.R. Detecting spot price forecasts in futures prices. *J. Bus.* **1986**, *59*, S39–S54. [CrossRef]

117. Kim, Y.S.;Walls, L.A.; Krafft, P.; Hullman, J. A Bayesian Cognition Approach to Improve Data Visualization. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, UK, 4–9 May 2019; p. 682.

118. Vul, E.; Goodman, N.; Griffiths, T.L.; Tenenbaum, J.B. One and done? Optimal decisions from very few samples. *Cogn. Sci.* **2014**, *38*, 599–637. [CrossRef] [PubMed]

119. Deshpande, S. Brexit Myth on FTSE and DAX Companies: A Review. Available at SSRN 3517139. 2020. Available online: https://www.researchgate.net/profile/Shubhada-Deshpande-2/publication/338502066_Brexit_Myth_on_FTSE_and_DAX_Companies_A_Review/links/5e183584a6fdcc2837662070/Brexit-Myth-on-FTSE-and-DAX-Companies-A-Review.pdf (accessed on 10 May 2021)

120. Cox, J.; Griffith, T. Political Uncertainty and Market Liquidity: Evidence from the Brexit Referendum and the 2016 US Presidential Election. Available at SSRN 3092335. 2018. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092335 (accessed on 10 May 2021)