The Wisdom and Manipulability of Threads

ROBIN ENGELHARDT, JACOB STÆRK-ØSTERGAARD and VINCENT FELLA HENDRICKS,
University of Copenhagen

1. INTRODUCTION

Social information in the form of opinions and judgments by other people is sampled sequentially. We read the news, hear rumors, listen to debates on TV, and flip through comments on social media platforms and blogs. These activities inform us and influence our decisions, but researchers still debate the conditions under which these types of social information help us make better decisions [Woolley et al. 2010; Gürçay et al. 2015; Becker et al. 2017; Jayles et al. 2017], lead us astray [Lorenz et al. 2011; Minson and Mueller 2012; Le Mens et al. 2018], or just make us confused at a higher level [Salganik et al. 2006].

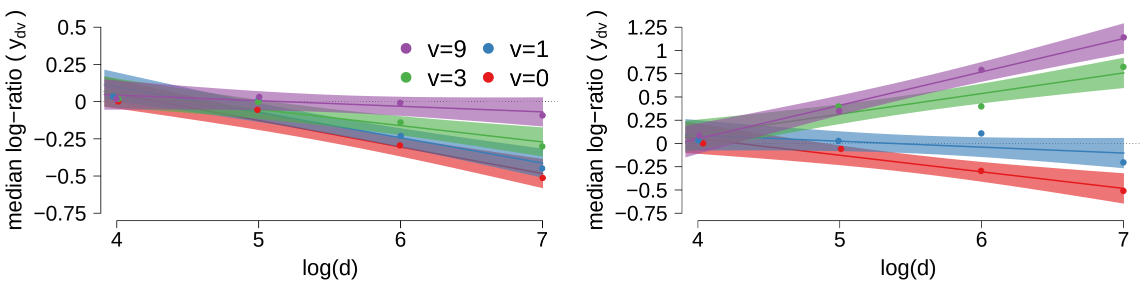
Collective estimates of a diverse group of people can outperform the majority of its members because any random confusion at the individual level is likely to average out and let the most accurate estimate prevail [Galton 1907; Surowiecki 2005]. Then again, confusion is not always randomly scattered around the truth. Systematic biases in individual perception may create measurable disruptions in the wisdom of crowds [Izard and Dehaene 2008; Kao et al. 2018]. Social information can add to those biases and create echo chambers, bandwagoning, herd behavior [Bikhchandani et al. 1992; Bakshy et al. 2015; Banerjee 1992], and various belief misattributions [Katz et al. 1931; Ross et al. 1977; Lee et al. 2019], which uphold harmful social practices despite being rejected by a majority of people. Social information may also have been intentionally filtered or manipulated in various ways, for instance through group pressure, algorithmic filtering [Pariser 2011], false cues [Muchnik et al. 2013; Hanson and Putler 1996], or simply by plain misinformation [Hendricks and Vestergaard 2018], often with highly detrimental consequences for our economy and our health.

Observational data of decision-making processes is acutely sensitive to the social context in which people find themselves. Thus, researchers find it difficult to separate observational data into its social and individual components. How may we know how much weight an individual puts on her own 'independent' estimate relative to the weight put on the estimates by others? Randomized experimental studies have attempted to solve this problem by first letting participants make a magnitude estimate of an object without social information (*ex ante*), and subsequently ask them to revise their estimate after having received information about other people's estimates of that object (*ex post*) [Becker et al. 2017; Jayles et al. 2017; Lorenz et al. 2011; Mavrodiev et al. 2013]. This double elicitation paradigm presumes that people change their mind because of the social information they have received. Other studies, however, have shown that people routinely can change their mind all by themselves, and that it may be more correct to assume an 'inner crowd' in the sense that people sample randomly from a probability distribution in their own mind [Vul and Pashler 2008; Herzog and Hertwig 2014]. Such a psychological mechanism (and others such as hedging strategies due to anticipated regrets [Bell 1982], and/or disappointments [Loomes and Sugden 1986]) make it difficult to differentiate accurately between 'inner' samples and 'outer' influences, and it may be desirable to develop an alternative model framework that is able infer the extend of individual bias and social influence from a single task.

We propose to use probabilistic gaussian mixture models (GMMs) since they have two properties that are highly valuable in the context of free response elicitation tasks. First, GMMs are comprised of several Gaussians which fit well to the often highly right-skewed and long-tailed distributions of estimates. Second, GMMs associate an uncertainty measure to each data point, telling us how much an estimate may be associated to a sub-population of people who are influenced in a certain way by the social information, without having any prior identity information about sub-populations in the data set.

2. THREAD MEDIANS

Using a cascade design that emulates succesive decision-making, we investigate experimentally the collective accuracy and manipulability of threads. The data was collected from 7,814 participants playing a dot-guessing game [Horton 2010] on Amazon Mechanical Turk, where each participant successively estimated the number of visible dots $d \in \{55, 148, 403, 1097\}$ in various images, while seeing $v \in \{0, 1, 3, 9\}$ preceding estimates (historical threads). An additional 3,934 participants were shown the same images while seeing the $v \in \{1, 3, 9\}$ *highest* estimates made so far (manipulated threads).



(a) Historical threads. (b) Manipulated threads.

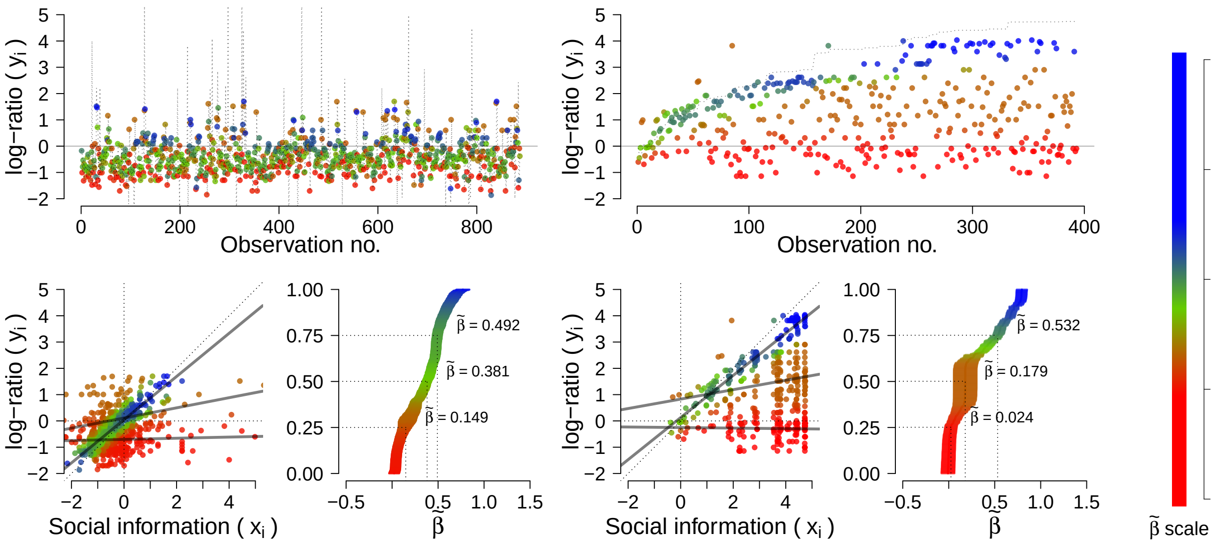
Fig. 1: Relationship between median log-ratio y_{dv} and $\log(d)$ with 95% confidence bounds (shaded areas). The colors represent the four settings on number of visible estimates. There is a clear relation between number of dots and number of visible estimates in both thread types, but it is more pronounced for the manipulated series. The red lines corresponding to the control groups with $v = 0$ are the same in both plots.

We compare the log-ratio of the individual thread medians, $y_{dv} = \log(M_{(d,v)}/d)$, using a linear normal model to quantify differences between threads in terms of $\log(d)$ and v, the latter as a categorical variable. Figure 1a shows that the collective performance for historical threads declines significantly with increasing task difficulty (higher d), but improves when the social information is substantial (high v). In contrast to [Lorenz et al. 2011] and in accordance with [Gürçay et al. 2015; Becker et al. 2017], this finding supports the claim that crowds indeed may become wise under (pristine) social influence. It should be noted, however, that the overlapping confidence intervals reveal where thread performances are comparable. The combination of the negative effects of task difficulty and the positive effects of social information are only discernible in situations where people have hard problems to solve and at the same time have plenty of social information available.

For the manipulated series, Fig.1b, the manipulation gives a large positive bias for $v = 3, 9$ which increases with d, implying that when a task becomes more demanding, the amount of manipulated social information has a detrimental impact on thread performance.

3. SOCIAL INFLUENCE

Figure 2 displays the effect of social information, obtained by fitting a GMM to each thread using the geometric mean of visible estimates as a proxy for the available social information. The effect of social information on each participant is given by $\tilde{\beta}$, the weighted average of the individual states of the GMM, where a high value ($\tilde{\beta} > 0.6$, blue color) implies a large effect, a low value ($\tilde{\beta} \approx 0$, red color) implies little or no effect and a medium value ($\tilde{\beta} \approx 0.4$, green color) suggests a compromise between the two extremes.



(a) Historical thread with $d = 1097$ and $v = 1$ (b) Manipulated thread with $d = 1097$ and $v = 9$

Fig. 2: The left hand side shows three plots of a historical thread with $d = 1097$ and $v = 1$, and the right hand side shows three plots of a manipulated thread with $d = 1097$ and $v = 9$. Top plots shows the log-ratio estimates over time (observation no.), with the geometric mean of the social information shown by a dotted line. Bottom left plots show the log-ratio of the estimates as a function of the log-ratio of the social information, indicating how differently participants use their social information. Bottom right plots show the cumulative distribution of individual $\tilde{\beta}$'s with 95% intervals derived from the fitted models and added interquartile values of $\tilde{\beta}$.

Figure 2a shows results for a historical thread where estimates are split among keepers (red), compromisers (green) and adopters (blue) throughout the thread. Figure 2b in contrast shows the results for a manipulated thread, where the same three groups are clearly visible, but the evolutionary trend is completely different compared to the historical thread. Contrasting Figures 2a and 2b clearly reveals different dynamics of threads in terms of manipulation or not. In general, our experiments show distinct distributions of overlapping groups (keepers, compromisers and adopters) in each thread, and in the case of manipulation suggest a substantial split into enthusiastic adopters versus skeptic keepers. The GMM framework is also applied to independent data from [Jayles et al. 2017] allowing similar insights into qualitatively different tasks.

REFERENCES

Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

Abhijit V Banerjee. 1992. A simple model of herd behavior. *Quart J Econ* 107, 3 (1992), 797–817.

Joshua Becker, Devon Brackbill, and Damon Centola. 2017. Network dynamics of social influence in the wisdom of crowds. *P Natl Acad Sci Usa* 114, 26 (2017), E5070–E5076.

David E Bell. 1982. Regret in decision making under uncertainty. *Operations research* 30, 5 (1982), 961–981.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *J Polit Economy* 100, 5 (1992), 992–1026.

Francis Galton. 1907. Vox populi (the wisdom of crowds). *Cah Rev The* 75, 7 (1907), 450–451.

Burcu Gürçay, Barbara A Mellers, and Jonathan Baron. 2015. The power of social influence on estimation accuracy. *J Behav Decis Making* 28, 3 (2015), 250–261.

Ward A Hanson and Daniel S Putler. 1996. Hits and misses: Herd behavior and online product popularity. *Market Lett* 7, 4 (1996), 297–305.

Vincent F Hendricks and Mads Vestergaard. 2018. *Reality Lost: Markets of Attention, Misinformation and Manipulation*. Springer.

Stefan M Herzog and Ralph Hertwig. 2014. Harnessing the wisdom of the inner crowd. *Trends in cognitive sciences* 18, 10 (2014), 504–506.

John J Horton. 2010. The Dot-Guessing Game: A 'Fruit Fly'for Human Computation Research. *SSRN 1600372* (2010). DOI:http://dx.doi.org/10.2139/ssrn.1600372

Véronique Izard and Stanislas Dehaene. 2008. Calibrating the mental number line. *Cognition* 106, 3 (2008), 1221–1247.

Bertrand Jayles, Hye-rin Kim, Ramón Escobedo, Stéphane Cezera, Adrien Blanchet, Tatsuya Kameda, Clément Sire, and Guy Theraulaz. 2017. How social information can improve estimation accuracy in human groups. *P Natl Acad Sci Usa* 114, 47 (2017), 12620–12625.

Albert B Kao, Andrew M Berdahl, Andrew T Hartnett, Matthew J Lutz, Joseph B Bak-Coleman, Christos C Ioannou, Xingli Giam, and Iain D Couzin. 2018. Counteracting estimation bias and social influence to improve the wisdom of crowds. *J R Soc Interface* 15, 141 (2018), 20180130.

Daniel Katz, Floyd Henry Allport, and Margaret Babcock Jenness. 1931. *Students' attitudes; a report of the Syracuse University reaction study.* Technical Report. Oxford.

Gaël Le Mens, Balázs Kovács, Judith Avrahami, and Yaakov Kareev. 2018. How endogenous crowd formation undermines the wisdom of the crowd in online ratings. *Psychol Sci* 29, 9 (2018), 1475–1490.

Eun Lee, Fariba Karimi, Claudia Wagner, Hang-Hyun Jo, Markus Strohmaier, and Mirta Galesic. 2019. Homophily and minority-group size explain perception biases in social networks. *Nature human behaviour* (2019), 1–10.

Graham Loomes and Robert Sugden. 1986. Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies* 53, 2 (1986), 271–282.

Jan Lorenz, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. How social influence can undermine the wisdom of crowd effect. *P Natl Acad Sci Usa* 108, 22 (2011), 9020–9025.

Pavlin Mavrodiev, Claudio J Tessone, and Frank Schweitzer. 2013. Quantifying the effects of social influence. *Sci Rep* 3 (2013), 1360.

Julia A Minson and Jennifer S Mueller. 2012. The cost of collaboration: Why joint decision making exacerbates rejection of outside information. *Psychol Sci* 23, 3 (2012), 219–224.

Lev Muchnik, Sinan Aral, and Sean J Taylor. 2013. Social influence bias: A randomized experiment. *Science* 341, 6146 (2013), 647–651.

Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Lee Ross, David Greene, and Pamela House. 1977. The "false consensus effect": An egocentric bias in social perception and attribution processes. *J Exp Soc Psychol* 13, 3 (1977), 279–301.

Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311, 5762 (2006), 854–856.

James Surowiecki. 2005. *The wisdom of crowds*. Anchor.

Edward Vul and Harold Pashler. 2008. Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science* 19, 7 (2008), 645–647.

Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science* 330, 6004 (2010), 686–688.