The Influence of Early Respondents: Information Cascade Effects in Online Event Scheduling

Daniel M. Romero
School of Information
University of Michigan
drom@umich.edu

Katharina Reinecke
Computer Science &
Engineering, DUB group
University of Washington
reinecke@cs.washington.edu

Lionel P. Robert Jr.
School of Information
University of Michigan
lprobert@umich.edu

ABSTRACT

Sequential group decision-making processes, such as online event scheduling, can be subject to social influence if the decisions involve individuals' subjective preferences and values. Indeed, prior work has shown that scheduling polls that allow respondents to see others' answers are more likely to succeed than polls that hide other responses, suggesting the impact of social influence and coordination. In this paper, we investigate whether this difference is due to information cascade effects in which later respondents adopt the decisions of earlier respondents. Analyzing more than 1.3 million Doodle polls, we found evidence that cascading effects take place during event scheduling, and in particular, that early respondents have a larger influence on the outcome of a poll than people who come late. Drawing on simulations of an event scheduling model, we compare possible interventions to mitigate this bias and show that we can optimize the success of polls by hiding the responses of a small percentage of low availability respondents.

Keywords

Decision-making; Doodle; Information cascade; Herding behavior; Social influence

1. INTRODUCTION

Groups are often preferred as decision makers over individuals because, collectively, they have access to more information [28, 30, 32]. However, group decision-making processes can be subject to social influence, which can undermine their ability to make unbiased decisions [26]. Social influence occurs when group members simply conform to or adopt the behavior of other members [25]. Thus, a final decision might not be the result of its members' unbiased inputs, but instead the result of a biased decision-making process derived from information cascades [3, 4, 5, 24], which can distort the outcome.

Online event scheduling is a type of group decision-making process, which supports groups in finding a mutually agreeable time. While event scheduling could be expected to be less prone to social influence when compared to other opinion-based processes such as voting, decisions in event scheduling are not completely binary [23]. Prior research suggests that respondents of event scheduling polls influence each other's responses. Open polls, in which respondents can see others' availability, result in a higher number of mutually agreeable times than hidden polls, in which others' responses are obscured [27] — a finding that suggests the presence of social influence. Depending on whether and how early respondents influence later respondents, this effect could have a negative or positive impact on the likelihood that the group finds a time that accommodates a large number of the group members.

To analyze the effect of social influence in event scheduling via polls, this work investigates three questions: (1) Do early respondents have a larger impact on the overall success of a poll compared to late respondents? If so, this would suggest that social influence in event scheduling generates cascading effects that form early on in the polling process. (2) Do preceding members influence the listed availability of proceeding group members and does the effect diminish as users are further apart in order of arrival to the poll? Understanding this would help further explain the mechanics of local social influence that could result in global cascading effects. (3) What interventions can be used by designers of event scheduling services like Doodle to mitigate the negative impact of social influence and exploit its positive impact?

To answer these questions, we examined data from more than 1.3 million polls generated by users of the online event-scheduling site Doodle.[1] Our findings show that early respondents indeed have a larger influence on the outcome of a poll than people who came later. This finding does not hold for hidden polls, showing that the effect is due to social influence. We also find that respondents are influenced the most by others who posted just before them. This indicates that preceding respondents have a larger influence on directly following respondents. Finally, based on the previous results, we explore an intervention in the context of a simulation based on an event scheduling model. The intervention aims to minimize the negative effects of social influence and increase the success of event scheduling polls. We find that we can eliminate the negative effects of social influence by hiding low availability responses from less than 5% of the respondents, and we can optimize poll success through the positive effects of social influence by hiding 35% to 65% of low availability responses.

The results of our analysis contribute to the literature by 1) examining how social influence materializes in event scheduling, 2) identifying its negative impact on the success of event scheduling polls, and 3) presenting an intervention that minimizes this negative impact on the success of event scheduling polls. In doing so, our results inform the design of future event scheduling systems by

WSDM 2017, February 06-10, 2017, Cambridge, United Kingdom

DOI: http://dx.doi.org/10.1145/3018661.3018725

[1]http://www.doodle.com

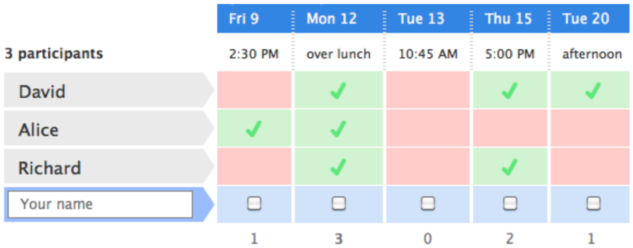


Figure 1: An example poll on the online event scheduling site Doodle.

showing how to limit the negative effects of social influence and exploit its positive effects for more successful event scheduling.

2. RELATED WORK

Social influence occurs when an individual's views, emotions, or actions are impacted by the views, emotions or actions of another individual [15]. Social influence has been observed in many domains including health behaviors [13], political participation [9], purchasing habits [19], collective decision-making (e.g. Wikipedia promotions) [22], idea generation [1, 2, 34], and online news voting [25, 36]. In particular, the first few people who voice their opinion or vote can determine the overall outcome [22, 25]. In other words, early respondents tend to influence the members who arrive later. This makes early arriving members the most influential in a sequential group-decision making process.

Group think, peer pressure, or the need to conform to others are all examples of social influence [18]. At the core of social influence is the underlying problem that groups fail to reach a decision based on the unique and independent information of its members [37]. Instead, groups reach decisions by having many members attempt to match their preferences to the already stated views of a few members [14].

Information cascades are particular instances of social influence [17]. They occur when individuals not only follow the actions of others, but do so despite holding private information that may contradict the actions taken by others [16, 21]. Information cascades have been observed in empirical and experimental settings [3, 4, 24] and a wide range of mathematical models and empirical studies have been developed and conducted to understand cascade dynamics [6, 7, 35, 11, 12, 38]. The existing models are very general and apply to any setting where information cascades may occur. We contribute to this work by extending previous cascade models such that they apply specifically to event scheduling.

Another related phenomenon that results from social influence is herd behavior [5]. Herd behavior occurs when individuals choose to behave in the same way as others in a crowd, but unlike information cascades they do not hold private information which contradicts the actions of others [33]. While information cascades and herd behavior are similar and often hard to tell apart, it is possible to distinguish between the two through experiments [10].

In the case of scheduling, there is some evidence that social influence impacts the way that users respond to polls. The evidence comes from the differences between open polls, where respondents can see each other's listed availabilities, and hidden polls, where respondents cannot see each other's listed availabilities. One study found that respondents list a wider availability in open polls than in hidden polls [39]. Open polls are also more successful when it comes to reaching consensus [27]. This suggests that respondents in open polls adjust their availability to align with those of their colleagues; hence, increasing the opportunity to reach consensus. It has also been observed that responses of later respondents are correlated with earlier responses and this correlation is higher in open polls [39]. This suggests that seeing prior responses affects the way respondents answer the poll.

This paper aims to build on prior work that has identified social influence in event scheduling by analyzing and quantifying the impact that social influence has on scheduling polls. We have the particular goals of investigating the extent to which poll participants adopt the decisions of earlier respondents, how long the effect of a response lasts on future respondents, whether early respondents have a disproportionate impact on the success of a poll, and to provide possible strategies that may allow us to mitigate any negative impact of social influence.

3. DOODLE DATASET

Doodle is a scheduling service that allows groups of people to find a mutually agreeable time. An initiator creates a Doodle poll that is sent to group members. The poll asks each group member to declare whether they are available for each of the time slot options originally chosen by the poll initiator (see Figure 1). After all members have responded to the poll, the initiator can determine which slot is the most appropriate to schedule the event. This is usually the slot that accommodates the most group members.

Doodle offers both open and hidden polls. In open polls, respondents are able to see the responses of other group members and the order in which they arrived before responding to the poll. In this case, respondents have an opportunity to get a clear view of the availability of all respondents who came before them. In hidden polls, respondents cannot see others' responses – they list their availability without knowing the availability of others. Doodle also supports "if need be" polls, in which individuals can additionally choose a third option to indicate that they could be available if nothing else works. Most polls do not have the "if need be" option, therefore we do not include them in our study.

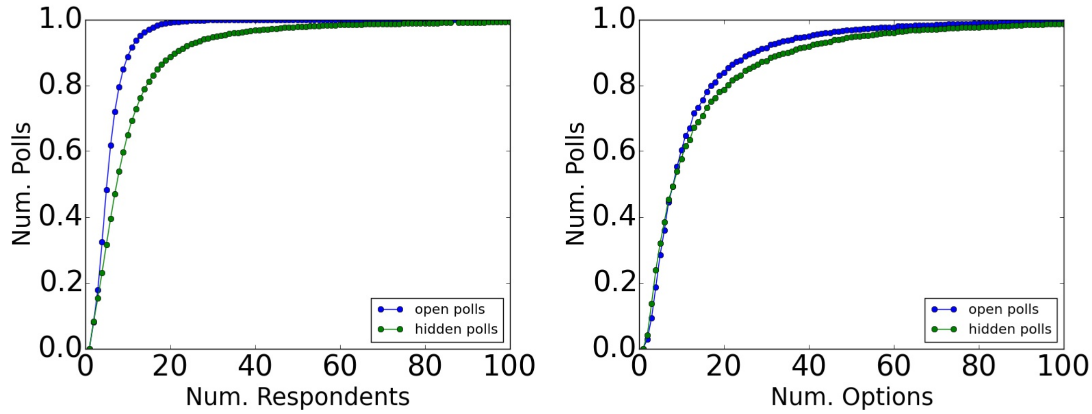
Our data[2] contains 1,303,941 anonymized open and 26,613 hidden yes/no polls. The users who generated these polls are in 190 different countries. Figure 2 shows the cumulative distribution function (CDF) of the number of responses and options of open and hidden polls. We observe that the number of respondents and options in the polls range from 2 to over 100. Also, open polls tend to have a smaller number of respondents and options than hidden polls.

4. INFORMATION CASCADE EFFECTS IN DOODLE POLLS

4.1 Poll success

We begin by defining and measuring the *success* of a Doodle poll. People use Doodle polls for different purposes. Sometimes they are trying to find a time slot that is available to all respondents and other times they are trying to find a slot that will accommodate the most people, even if there is not a time slot that works for all participants. Because we are not able to identify the goals that the participants of the poll had in mind, we use an intuitive measure of success: the percentage of participants who are available on the most popular time slot. This measure is exactly what participants are trying to maximize when they are looking for a time slot that works for most people. This also measures *how close* the participants are to finding a time that works for everyone. In either case, it is reasonable to assume that a poll where a high percentage of

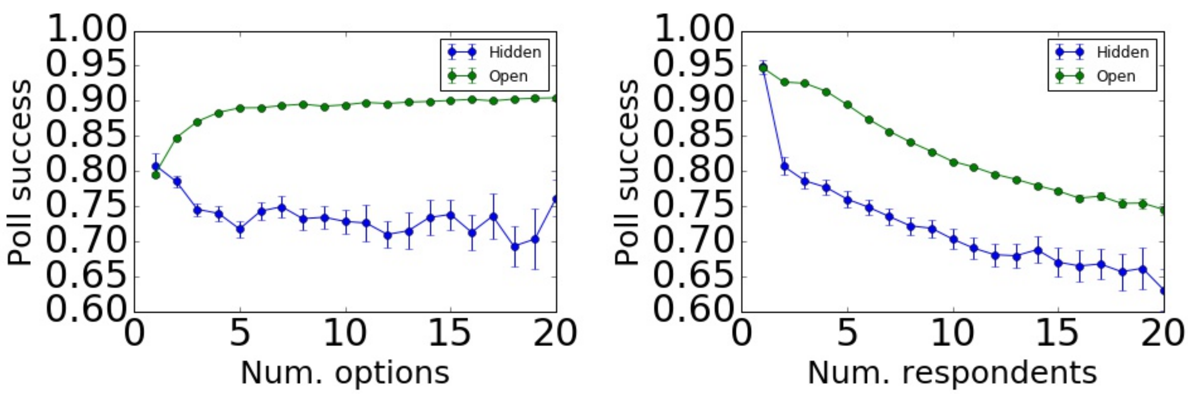
[2]This dataset was obtained in collaboration with Doodle.



(a) CDF of the number of respondents in open and hidden polls.

(b) CDF of number of options in open and hidden polls.

Figure 2: The cumulative distribution function (CDF) of the number of respondents and options of open and hidden polls.



(a) Success of polls by number of options.

(b) Success of polls by number of respondents.

Figure 3: Success of polls based on the number of respondents and options.

the participants are accommodated in some time slot is more successful than one where only a small percentage of participants are available in all time slots.

Note that our measure of success applies to the global success of the poll, not necessarily to the individual goals of the respondents. There are many considerations that respondents have in mind when they answer the poll, making their individual goals complex. For example, respondents may consider their own preferences, their relationship with other respondents, their perception of the importance of the event, and how the other respondents may perceive them based on their availability. Indeed, it has been observed that open polls on Doodle have higher response rates for highly popular and highly *unpopular* slots [39], which suggests that respondents attempt to coordinate with others and also attempt to appear flexible with their time by making themselves as available during slots that are unlikely to be chosen for the event. This highlights the complexity of individual goals, which we do not attempt to measure or model. Instead, we focus on the global success of a poll.

We formally define poll success as the following. Each poll has a number of people who respond to the poll sequentially and a number of options that respondents label as available or unavailable. For each poll p we define R_p and O_p as the number of respondents and options, respectively. We let $A_p(i, j)$ be an indicator function that indicates whether the i^{th} respondent is available for option j:

$$A_p(i, j) = \begin{cases} 1 & i^{th} \text{ respondent is available for option } j \\ 0 & \text{otherwise} \end{cases}$$

The success of poll p is $succ_p = \max_{j \in [1, O_p]} \frac{1}{R_p} \sum_{i=1}^{R_p} A_p(i, j)$. For each poll p, $succ_p$ is the percentage of respondents who are available for the option that accommodated the most respondents.

Throughout the paper we will measure how different characteristics of a poll affect its success. In this section we explore two basic properties that have a very clear effect on the poll's success – the number of options and the number of respondents. We will also compare how our measure differs when we consider open polls and hidden polls. We use the set of hidden polls as a control group of polls. While there could be systematic differences in the reasons that people choose to create open and hidden polls, the only way in which they differ on how they work on Doodle is that the responses on hidden polls are not public. Comparing effects in open and hidden polls allows us to identify effects that are due to respondents being able to observe each other's responses.

Figure 3 shows how the success of the polls changes with the number of respondents and options. We find that polls with a larger number of respondents tend to be less successful in both open and hidden polls than those with small number of respondents, showing that it is generally harder to accommodate a larger group of people. The number of options has a positive effect on the success of open polls, suggesting that when more time slots are available, more people can be accommodated. However, for hidden polls, the number of options has a slightly negative effect on success as the number of options increases from 1 to 5. Beyond 5, there is no relationship between number of options and success. This contrast between the effect of number of options in open and hidden polls suggests that in order to take advantage of a larger number of available slots, respondents need to be able to coordinate, and hence need to see each other's responses.

4.2 Impact of Early Respondents

One of our goals is to explore the role of early respondents on the overall success of the poll, and in particular the effect of early respondents with low availability. If respondents are simply reporting their availability, independently of the availability of other respondents, then we should observe no effect. However, if respondents are able to observe others' responses and take these observations into account when reporting their own availability, then having an early respondent with low availability could have a negative effect on the overall poll. When a new respondent observes that earlier respondents reported low availability, this removes social pressure from the new respondent to report a wide availability. This effect could cascade through the sequence of respondents, making each new respondent more likely to report a limited availability.

We measure the extent to which early respondents of poll p reported low or high availability in three ways. First, we identify when the respondent with the lowest availability came in the sequence of respondents – we let $loc\_min = \underset{i \in [1,R_p]}{\arg\min} \frac{1}{O_p} \sum_{j=1}^{O_p} A_p(i,j)$ be the location of the respondent with the lowest availability. Second, we identify when the respondent with the highest availability arrived – we let $loc\_max = \underset{i \in [1,R_p]}{\arg\max} \frac{1}{O_p} \sum_{j=1}^{O_p} A_p(i,j)$ be the location of the respondent with the highest availability. These first two measures capture the location of the extreme respondents. Third, we also measure the general trend of how the availability of the respondents changes as the respondents arrive to the poll. To do this, we let $a_p^i = \frac{1}{O_p} \sum_{j=1}^{O_p} A_p(i,j)$ be the fraction of options the i^{th} respondent reported as available. Then we compute the Spearman's rank correlation coefficient between the sequences $\{a_p^i\}_{i=1}^{R_p}$ and $\{i\}_{i=1}^{R_p}$. We let $trend_p$ be this correlation. Thus, $trend_p$ measures the extent to which the sequence of availability of the respondents tends to increase or decrease as they arrive to the poll. When $trend_p$ is positive (negative), the availability of the respondents tends to increase (decrease) as they arrive.

Figure 4 shows the relationship between the success of a poll ($succ_p$) and each one of the following variables in open polls: location of the respondent with lowest availability (loc_p^{min}), location of the respondent with highest availability (loc_p^{max}), and the trend in availability ($trend_p$). Since we know that the number of respondents have a high impact on $succ_p$, Figure 4 shows results separately for polls with 2-7 respondents. This accounts for about 80% of all polls. We observe that $succ_p$ generally increases with loc_p^{min} and decreases with loc_p^{max} and $trend_p$. The results consistently indicate that when early respondents report low availability, the overall success of the poll decreases.

In order to further investigate the relationship between loc_p^{min}, loc_p^{max}, and $trend_p$ and the success of the polls, independently of the number of respondents and options of the poll, we run OLS regressions of the form $succ_p = \beta_1 R_p + \beta_2 O_p + \beta_3 loc_p^{min} + \beta_4 loc_p^{max} + \beta_5 trend_p + \epsilon_p$. Table 1 shows the coefficients and p-values of each variable in open polls. We find that there is a significant relationship ($p < 0.01$) between all the independent variables and the success of open polls. Furthermore, the sign of the coefficients is consistent with the observations in Figure 4.

Our result suggests that early respondents influence the rest of the respondents. This assumes that when a respondent comes to the poll, she is aware of the responses of previous respondents. We also produce the plots in Figure 4 using hidden polls, in which respondents are unable to see other responses. Figure 5 shows the relationship between $succ_p$ and loc_p^{min}, loc_p^{max}, and $trend_p$ in hidden polls. The only effect we observe in hidden polls happens when loc_p^{min} and loc_p^{max} change from 1 to 2 and $succ_p$ increases slightly in both cases. Furthermore, we run the same OLS regressions using hidden polls. Table 2 shows the coefficients and p-values. We find that only the variables R_p and $trend_p$ remain significant ($p < 0.01$), suggesting that most of the effects observed in open polls disappear in hidden polls. This indicates that the effects we observed in Figure 4 and Table 1 depend on the ability of the respondents to see others' responses. This aligns with our hypothesis that respondents are influenced by the level of availability of previous respondents.

4.3 Impact of Nearby Respondents

In the previous section we found that the availability of early respondents and the increasing or decreasing trend in availability as respondents arrive to the poll have an impact on the overall success of the poll. One way to explain this effect is that as respondents arrive and observe a limited availability of previous respondents, they may feel less pressure to report a wide availability. But how long does the availability of a respondent continue to have an effect on future respondents? That is, how much does the availability of respondent i relate to the availability of respondent $i+k$? To answer this question we compare the availability reported by users who are near each other in the list of respondents.

Recall that $a_p^i = \frac{1}{O_p} \sum_{j=1}^{O_p} A_p(i,j)$ indicates the fraction of options for which the i^{th} respondent is available in poll p. For each poll p, we compute the mean availability of the respondents as $m_p = mean(a_p^i)$. We then define an indicator function, b_p^i, which indicates whether the i^{th} respondent has above or below average availability, as:

$$b_p(i,j) = \begin{cases} 1 & a_p^i \geq m_p \\ 0 & \text{otherwise} \end{cases}$$

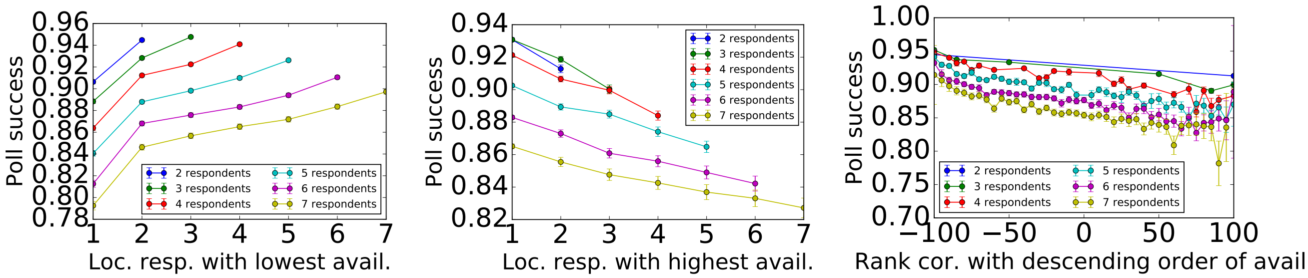
We compute the probability that two users who are k spots away from each other (that is, one user is the i^{th} respondent and the other users if the $(i+k)^{th}$ respondent) are both above or below average, or whether one is above average and the other is below average. We compute this probability separately for polls with $3, \ldots, 7$ respondents. More specifically, we let $c_p(i,j) = 1$ if $a_p^i = a_p^j$ and $c_p(i,j) = 0$ if $a_p^i \neq a_p^j$, and define $P_r^k = \frac{1}{(r-k)|S_r|} \sum_{p \in S_r} \sum_{i=1}^{r-k} c_p(i, i+k)$, where S_r is the set of polls with exactly r respondents. P_r^k corresponds

Independent Variable	Coefficient	p-value
Num. Respondents (R_p)	-0.0112	0.000
Num. Options (O_p)	0.0002	0.000
Loc. least available respondent (loc_p^{min})	0.0091	0.000
Loc. most available respondent (loc_p^{max})	-0.0039	0.000
Increasing availability trend ($trend_p$)	-0.0156	0.000

Table 1: OLS regressions for open polls $succ_p = \beta_1 R_p + \beta_2 O_p + \beta_3 loc_p^{min} + \beta_4 loc_p^{max} + \beta_5 trend_p + \epsilon_p$

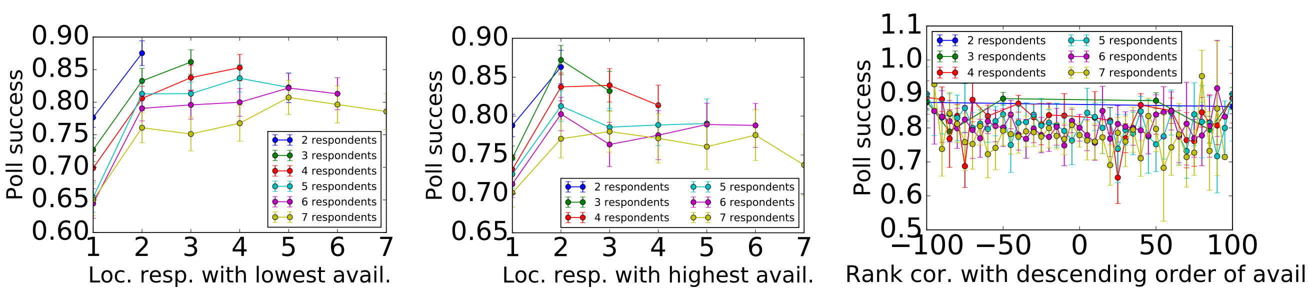
Independent Variable	Coefficient	p-value
Num. Respondents (R_p)	-0.0017	0.00
Num. Options (O_p)	0.0000	0.769
Loc. least available respondent (loc_p^{min})	0.0001	0.597
Loc. most available respondent (loc_p^{max})	-0.0008	0.018
Increasing availability trend ($trend_p$)	-0.0260	0.000

Table 2: OLS regressions for hidden polls $succ_p = \beta_1 R_p + \beta_2 O_p + \beta_3 loc_p^{min} + \beta_4 loc_p^{max} + \beta_5 trend_p + \epsilon_p$



(a) (b) (c)

Figure 4: Poll success vs. (a) *loc_min*, **(b)** *loc_max***(b), and (c)** *trend* **plotted separately for open polls with 1-7 respondents.**



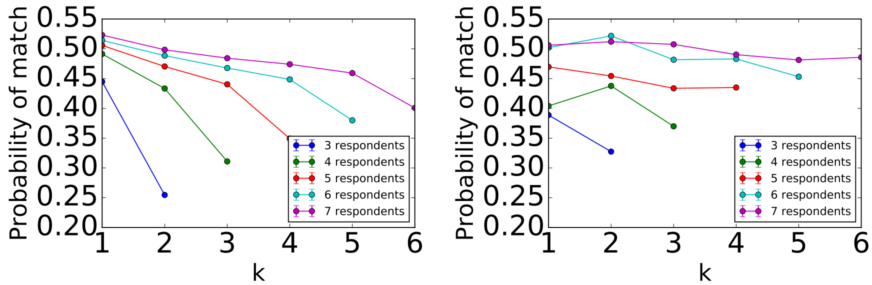
(a) (b) (c)

Figure 5: Poll success vs. (a) *loc_min*, **(b)** *loc_max***(b), and (c)** *trend* **plotted separately for hidden polls with 1-7 respondents.**

to the probability that users in polls with r respondents match in having above or below average availability when they are k users from each other in the list of respondents. Figures 6(a) and 6(b) show curves P_r^k vs k for $r = 3, \ldots, 7$ in open and hidden polls, respectively. We observe that P_r^k decreases with k in open polls, suggesting that users tend to match the availability of nearby users and the matching decreases as the gap between the users increases. In hidden polls, however, we observe that there is little or no relationship between P_r^k and k, showing that when users are not aware of each others responses, they do not match their availability.

The result suggests that the effect of early respondents on the success of the polls is possibly due to a mechanism where respondents influence other nearby respondents. When an early respondent reports low availability, this has a cascading effect over the rest of the poll and results in individuals reporting a low availability, making the poll less successful.

Zou et al. found that the availability of polls in Doodle tends to decrease as more respondents complete the poll [39], which could affect some of our measures. However, they found that the decreasing rate in availability is the same in open and hidden polls. Since our analysis is largely based on the differences in the dynamics between open and hidden polls, we conclude that the decreasing trend observed in [39] does not explain our findings.



(a) Open polls (b) Hidden polls

Figure 6: Probability of availability matching (P_r^k) vs. the lag between users (k) in open and hidden polls.

5. MODELING INFORMATION CASCADING BEHAVIOR IN EVENT SCHEDULING

In order to understand our empirical results further, we now develop a model of scheduling that explicitly takes into account the observation that early respondents tend to influence the responses of subsequent respondents. Our goal with this model is to simulate and compare potential interventions that a scheduling service such as Doodle can implement in order to alleviate some of the negative impact of this cascading behavior.

5.1 Model description

In the model, each poll has a fixed number of respondents r and a number of time slots s. We also fix a parameter k that denotes *how many* subsequent respondents each respondent will influence and a parameter $\alpha \in [0, 1]$ that denotes *how much* early respondents influence future respondents. Respondents arrive sequentially and declare whether they are available or unavailable for each of the slots. When deciding whether to report an option s as available or not, respondent i considers two things. First, i considers her private preference $pref_s^i \in [0, 1]$ for slot s. Here, we assume that rather than having binary preferences for time slots such as "busy" or "free", respondents have a continuous preference for each time slot that ranges from 0 (lowest preference) to 1 (highest preference).

Additionally, i considers the average fraction of slots that the previous k respondents marked as available, $prev_k^i$. As we observed in our empirical findings, our model assumes that respondents tend to correlate their availability with that of earlier respondents. Thus, respondent i reports s as available with probability $p_{s,k}^i = \alpha prev_k^i + (1 - \alpha) pref_s^i$. That is, respondent i puts a weight of α on the availability of previous respondents and a weight of $1 - \alpha$ to her own preference. The process continues until all respondents have declared their availability. Note that the parameter α controls how much respondents weight their own preferences compared to the availability of earlier respondents. When α is large, respondents weight the availability of earlier respondents higher than their own preferences, and vice versa when α is small. In other words, the parameters α and k control social influence.

We begin by asking whether social influence, which is represented by parameters α and k, has a positive or negative effect on the success of the polls. We run 50,000 simulations of the model with 10 respondents, 10 options[3], and various values of α and k. In the simulations each respondent i chooses her preference for slot s, $pref_s^i$, uniformly at random. We measure the success of the poll for each run as defined in the previous section. Intuitively, the effect of social influence should depend on the whether the initial set of respondents had high or low availability. However, we observe that overall, the success of the polls decreases with α and k. Figure 7(a) shows the relationship between success, α and k. In our model, social influence tends to have a slightly negative net impact on the success in scheduling. Given this finding, we now explore two potential interventions to mitigate the negative impact of social influence on scheduling.

5.2 Interventions

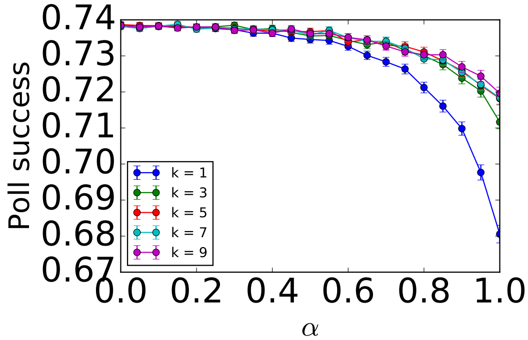
Since in our model social influence negatively impacts the success of polls, a natural strategy is to hide the responses of some users in order to prevent them from influencing the responses of others. In practice, hidden polls in Doodle have this feature but also prevent users from coordinating. For example, in open polls, respondents can shift their schedules in order to be available for a slot that appears to be popular among other respondents. This type of coordination can increase the success of a poll [27]. Hence, our goal is to explore interventions that hide some of the responses, but not necessarily all.

5.2.1 Hiding early respondents

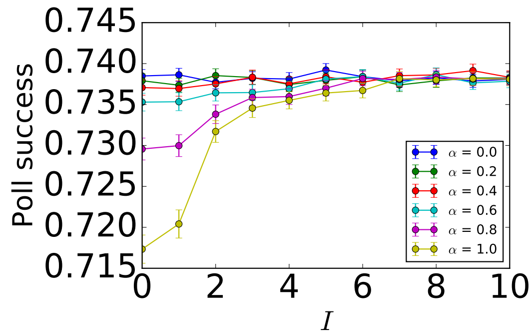
We begin by extending the scheduling model to include a parameter I, such that the responses from the first I users are hidden from all respondents. This would prevent cascades of low availability responses from forming early on in the poll. Note that this is effectively a hybrid between Doodle's open and hidden polls; since only a portion of the respondents are hidden, coordination among respondents who are not hidden is still possible. In the new version of the model, the responses of the initial I are not taking into account by the respondents when reporting their availability. This implies that the first I poll participants only take into account their own preferences and are not influenced by others. Also note that when I is the same as the number of respondents, the model is equivalent to a hidden Doodle poll, which also corresponds to $\alpha = 0$.

We simulate this version of the model with 10 respondents, 10 options, $k = 1$, and various values of α and I. Figure 7(b) shows the success of the simulated polls using different values of α and I. We observe that for all values of $\alpha > 0$, the success of the polls increases with I until it reaches the same success as when $\alpha = 0$. Hence, hiding an initial set of respondents increases the success of

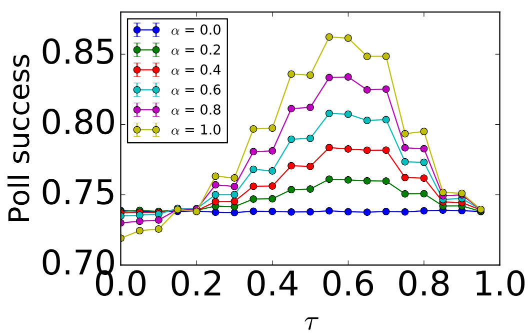
[3]All analyses presented in this section are based on 50,000 simulations with 10 respondents and 10 options. Simulations were also performed with different numbers of respondents and options (5 to 30) with the same results.



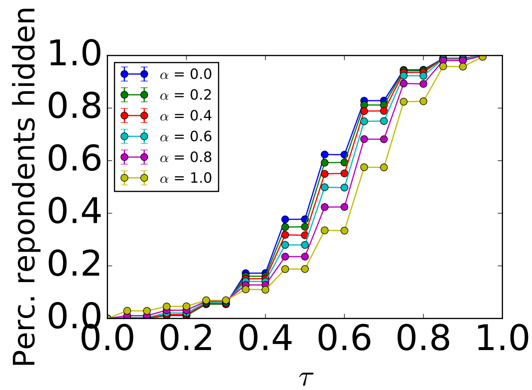
(a) Poll success vs. social influence parameters α and k. The net impact of social influence on poll success is negative.



(b) Poll success vs. number of initial hidden responses (I) for different values of α. Hiding initial users increases the success of the polls up to the baseline success when $\alpha = 0$ (no social influence).



(c) Poll success vs. threshold used when hiding responses with low availability (τ) for different values of α. Poll success initially increases with τ, but decreases as τ becomes large.



(d) Percentage of respondents hidden when hiding responses with low availability (τ) for different values of α.

Figure 7: Results from simulations of event scheduling model and interventions

the polls, but it never improves success above the baseline of $\alpha = 0$, which is equivalent to hiding all responses.

5.2.2 *Hiding low availability respondents*

Rather than hiding all the responses of all initial respondents, we now explore the effect of hiding responses of users with low availability. We extend our model by introducing a parameter $\tau \in [0, 1]$. When a respondent computes the average fraction of slots that the previous k respondents marked as available, $prev_k^i$, she will only take into account respondents who have at least a τ fraction of slots available. This effectively models an system where respondents with availability less than τ are hidden from other respondents.

We simulate the new model with 10 respondents, 10 options, $k = 1$, and various values of α and τ. Figure 7(c) shows that poll success initially increases as τ increases, which is the effect of hiding respondents with very low availability. However, as τ continues to increase, the success of the polls eventually begins decreasing. This is because when τ is too large, we are hiding respondents even when they have high availability, which would otherwise influence other respondents to increase their availability. Overall, we find that hiding respondents with low availability is an effective strategy as long as the threshold for hiding them is not too high.

Comparing the two strategies, we observe that hiding only those users with very low availability can be much more effective than hiding an initial set of respondents. For instance, even for very small values of τ such as $\tau = 0.15$, the success of the polls exceeds the baseline of $\alpha = 0$. Figure 7(d) shows the number of hidden responses vs. τ. When $\tau = 0.15$, less than 5% of the responses are hidden. This suggests that targeting a *small* number of responses to hide can be a very effective strategy to mitigate the negative effects of social influence. The intervention achieves the highest success when $\tau = 0.6$, which corresponds to hiding around 35% to 65% of the responses. The effects of coordination are not considered in the model, but note that coordination would only have an effect when at least some of the responses are not hidden. When $\tau = 0.6$, many responses are still visible, which would allow respondents to coordinate and potentially increase the success of the poll even more.

6. DISCUSSION

In this research, we sought to understand how social influence takes place and how it relates to the success of event scheduling.To do so, we examined more than 1.3 million polls generated by Doodle users. Results from our analysis provide three overarching findings:

1. Early poll respondents are most influential in determining the

success of polls. This suggests that respondents who come later engage in information cascades that ultimately limit the options for the following respondents. As a consequence, the availability of early respondents is vital to the success of the polls.

2. Respondents are influenced the most by their preceding neighbor. The availability of those before the preceding respondent still have some influence, but its impact diminishes rapidly. This localized influence can give rise to cascading effects that have a global impact on the poll.

3. Based on simulations of our event scheduling model, the likelihood of finding a mutually agreeable time increases when a small number of respondents with the lowest availability are hidden.

We now discuss these findings and their implications.

6.1 Early Respondents are Most Influential

Our findings suggest that there is strong social influence in event scheduling. This extends the work of Reinecke et al. [27], who found that open polls are more likely to result in mutually agreeable times than hidden polls and suggested that people must actively change their availability in order to find mutually agreeable times. We provide empirical evidence that information cascade effects are part of the process with the availability of early respondents greatly driving the success of polls.

If a respondent's availability is subject to social influence, event scheduling becomes a subjective choice process with people actively changing their availability to match or mismatch those of others. This can have positive effects, such as when respondents provide more availability than they actually have in order to find a mutually agreeable time. It can also have negative effects on the success of a poll, as our analysis showed, if early respondents indicate a low availability and others follow. In either case, event scheduling is no longer an objective choice-process that is based on facts (i.e., "my calendar shows that I am free at this time"), and its success is highly dependent on group dynamics and participants' willingness to place the group's goal above their personal interests. Hence, if a poll initiator wants to ensure that poll participants provide their truthful availability, the use of calendar systems, such as Microsoft Outlook or Google Calendar, that allow users to match previously entered availabilities might be a better option – assuming participants are willing to disclose this information (see also [20] for a discussion on privacy in scheduling).

6.2 Choices are Most Influenced by the Preceding Neighbor

The results of our analysis showed that people craft their availability to match those of immediately preceding poll respondents. A possible explanation for this is that the availability of a respondent's nearest preceding neighbor is perceived as an accurate depiction of the group's overall availability. As such, basing their availability on the information of the nearest preceding neighbor minimizes their effort. Related to this, they may have noticed a funneling effect where availability decreases with the least available options presented by the last respondent. The availability of their nearest preceding neighbor would then be a good representation of the options that are likely going to be successful.

An alternative explanation is the recency effect, which suggests that individuals tend to remember information presented at the end of a list best [8]. If this is true, then the location of the nearest neighbors might only heighten this effect; The information presented by the nearest preceding neighbor is more salient because of its location to the incoming poll participant's space provided to them to list their availability. In other words, respondents assume that their nearest preceding neighbor have taken into account the availability of others. Therefore, they may believe that matching with their nearest neighbor represents the best chance of achieving a successful poll.

While crafting their availability to match others can lead to a higher likelihood of poll success, we would want to discourage participants from listing only a small number of available times since this would decrease the number of participants who are accommodated at the most popular time slot.

6.3 Hiding Low-Availability Respondents Increases Success

There are several ways of mitigating the negative effect of social influence on the success of event scheduling. One possibility would be a design intervention that makes low-availability respondents aware of the negative influence of their choices in the hope that they will alter their availability. While raising awareness might convince some participants to edit their choices, such an intervention could be perceived as questioning the truthfulness of their availability.

Another option would be to shuffle responses in order to alleviate the funneling effect in which availability decreases with the number of poll respondents. Arriving poll participants may find it more difficult to see a pattern of funneling when the availability of respondents are presented randomly rather than sequentially. This, in turn, may reduce the bias associated with the nearest neighbor. While we believe that this could reduce people's tendency to be influenced by the availability of their immediately preceding neighbor, it is unlikely to solve the problem related to the disproportionate influence of the first few poll participants.

A third, more promising possibility is to hide responses that could potentially negatively influence proceeding poll participants. In our simulations, we showed that the likelihood of finding a mutually agreeable time increases when a small number of respondents with the lowest availability are hidden. The more respondents are hidden, the higher the likelihood of poll success, up to a point. We found that we can optimize poll success by hiding 35-50% of respondents with the lowest availability. We note that it's important to consider the specific goals of each poll before deciding if hiding low availability responses is likely to generate better results for the group. For example, in some polls the attendance of only a subset of individuals is required. In this case, it is possible that hiding the availability of these individuals would not yield a better outcome. Ultimately, the creator of the poll should be given the option of hiding all responses (hidden polls), no responses (open polls), or a subset of low availability of responses as we suggest in this paper.

The results of our simulation have implications for reducing social influence in other group-decision making processes. For example, social influence effects have been identified as a problem in online voting sites [25], with people often being biased by early responses. The results of our model indicate that it is possible to reduce this bias by hiding a given number of extreme responses. Some social media sites that include a voting system already temporarily hide the initial votes of content before users have consumed the content to avoid cascades from occurring [25]. Our findings suggests that this technique can be improved by hiding extreme prior votes or responses from arriving individuals before they cast their vote. Only once members have voted would the popularity of responses become visible. This could allow individuals to see a diversity of views while reducing the bias associated with social influence.

7. LIMITATIONS

We made several assumptions in our analysis that need to be taken into account when interpreting the results. First, our measure of success may not match the goals of all Doodle users. We operationalized success as a percentage of participants who are available on the most popular time slot. This assumes that Doodle users are always trying to find a slot that accommodates as many people as possible. However, in some cases, people may be attempting to find a slot that accommodates all respondents. In this case, our measure is a proxy for how close the poll was from being successful rather than a quantification of how successful the poll was.

Second, we assume that respondents are independent of each other. In doing so, we ignore the relationships between respondents. However, these relationships may impact the way respondents declare their availability. For example, some respondents may exert high influence over other respondents' schedules such as the case when people are scheduling a meeting with their boss or other people above them in a hierarchy. While previous research has highlighted the role of social hierarchy in determining individual availability [23], our data does not include information that would allow us to consider such relationships.

Third, hidden polls might be used for different purposes than open polls. To the degree that this is true, our comparison between hidden and open polls may over-attribute their differences to being able to see the listed times of others. Our data does not include such differences between open and hidden polls for reasons of anonymity, but a possible difference between hidden and open polls should be investigated in the future.

Finally, we do not account for coordination or a group member's intention to meet in our model. Coordination reflects the degree to which group members are able to organize and synchronize their actions [29, 31]. Respondents may actively coordinate their schedules in an attempt to find a time that best suits everyone. The effects of coordination could explain why open polls are generally more successful than hidden polls. We also did not include members' intention to make time available to meet. Intentions can reflect the degree to which someone is motivated to participate in the meeting[34]. It is possible that some respondents intentionally sought to undermine the scheduling process in order to avoid meeting. In the future, we plan to extend our current model to take this into account.

Ultimately, experiments and real interventions are necessary to establish the presence of social influence in scheduling and the effectiveness of the interventions. However, our results provide initial evidence that cascades occur in scheduling and suggest possible interventions that have the potential of making scheduling services more efficient.

8. CONCLUSION

This work provides evidence of information cascade effects in online event scheduling, showing that the success of a poll depends on the availability of early respondents and that participants are particularly influenced by the availability of their directly preceding neighbor. In many cases, this can mitigate the success of a poll (i.e., finding a mutually agreeable time). We therefore proposed a set of interventions that can help alleviate this bias and show that we can optimize the chances to find a mutually agreeable time. We hope that our findings help inform the design of future event scheduling systems and enable groups to more effectively arrange a time to meet.

9. ACKNOWLEDGMENTS

We thank Joey Hsiao, Danaja Maldeniya, Carol Moser, and Sarita Yardi Schoenebeck for their useful feedback on earlier drafts of this paper. This research was partly supported by the National Science Foundation under Grant No. IIS-1617820.

10. REFERENCES

[1] O. Alnuaimi, L. Robert, and L. Maruping. Social loafing in brainstorming cmc teams: The role of moral disengagement. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, pages 1–9. IEEE, 2009.

[2] O. A. Alnuaimi, L. P. Robert, and L. M. Maruping. Team size, dispersion, and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems*, 27(1):203–230, 2010.

[3] L. R. Anderson and C. A. Holt. Classroom games: Information cascades. *The Journal of Economic Perspectives*, 10(4):187–193, 1996.

[4] L. R. Anderson and C. A. Holt. Information cascades in the laboratory. *The American economic review*, pages 847–862, 1997.

[5] M. Baddeley. Herding, social influence and economic decision-making: socio-psychological and neuroscientific analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538):281–290, 2010.

[6] A. V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, pages 797–817, 1992.

[7] S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, pages 992–1026, 1992.

[8] R. A. Bjork and W. B. Whitten. Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, 6(2):173–189, 1974.

[9] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

[10] B. Çelen and S. Kariv. Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review*, pages 484–498, 2004.

[11] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pages 925–936. ACM, 2014.

[12] J. Cheng, L. A. Adamic, J. M. Kleinberg, and J. Leskovec. Do cascades recur? In *Proceedings of the 25th International Conference on World Wide Web*, pages 671–681. ACM, 2016.

[13] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[14] R. Cialdini and N. Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621, 2004.

[15] R. B. Cialdini and M. R. Trost. Social influence: Social norms, conformity, and compliance. In D. T. Gilbert, S. T. Fiske, and G. Lindzey, editors, *The handbook of social psychology*, volume 2, pages 151–192. McGraw-Hill, Boston, 1998.

[16] M. Eftekhar, Y. Ganjali, and N. Koudas. Information cascade at group scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 401–409. ACM, 2013.

[17] A. Guille, H. Hacid, C. Favre, , and D. A. Zighed. Information diffusion in online social networks: A survey. In *ACM SIGMOD Record*, pages 17–28. ACM, 2013.

[18] K. Herbert. Compliance, identification, and internalization: Three processes of attitude change. *Journal of conflict resolution*, pages 51–60, 1958.

[19] S. Hill, F. Provost, and C. Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, pages 256–276, 2006.

[20] B. Kellermann and R. Bohme. Privacy-enhanced event scheduling. In *Computational Science and Engineering (CSE'09)*, volume 3, pages 52–59. IEEE, 2009.

[21] D. E. J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[22] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg. Governance in social media: A case study of the wikipedia promotion process. In *ICWSM*, 2010.

[23] M. Mazmanian and I. Erickson. The product of availability: understanding the economic underpinnings of constant connectivity. In *ACM conference on Human Factors in Computing Systems*, pages 763–772. ACM, 2014.

[24] S. Milgram, L. Bickman, and L. Berkowitz. Note on the drawing power of crowds of different size. *Journal of personality and social psychology*, 13(2):79, 1969.

[25] L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

[26] J. Newell, L. Maruping, C. Riemenschneider, and L. Robert. Leveraging e-identities: The impact of percieved diversity on team social integration and performance. *ICIS 2008 Proceedings*, page 46, 2008.

[27] K. Reinecke, M. K. Nguyen, A. Bernstein, M. Näf, and K. Z. Gajos. Doodle around the world: Online scheduling behavior reflects cultural differences in time perception and group decision-making. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 45–54. ACM, 2013.

[28] L. Robert and D. M. Romero. Crowd size, diversity and performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1379–1382. ACM, 2015.

[29] L. P. Robert. Far but near or near but far?: The effects of perceived distance on the relationship between geographic dispersion and perceived diversity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2461–2473. ACM, 2016.

[30] L. P. Robert and D. M. Romero. The influence of diversity and experience on the effects of crowd size. *Journal of the Association for Information Science and Technology*, 2016.

[31] L. P. Robert Jr. Monitoring and trust in virtual teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 245–259. ACM, 2016.

[32] L. P. Robert Jr, A. R. Dennis, and M. K. Ahuja. Social capital and knowledge integration in digitally enabled teams. *Information Systems Research*, 19(3):314–334, 2008.

[33] L. Smith and P. Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

[34] S.-S. Srinivasan, L. M. Maruping, and L. P. Robert. Idea generation in technology-supported teams: A multilevel motivational perspective. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pages 247–256. IEEE, 2012.

[35] I. Welch. Sequential sales, learning, and cascades. *The Journal of finance*, 47(2):695–732, 1992.

[36] T. Weninger, T. J. Johnston, and M. Glenski. Random voting effects in social-digital spaces: A case study of reddit post submissions. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 293–297. ACM, 2015.

[37] P. Zarnoth and J. A. Sniezek. The social influence of confidence in group decision making. *Journal of Experimental Social Psychology*, 33(4):345–366, 1997.

[38] Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1513–1522. ACM, 2015.

[39] J. Zou, R. Meif, and D. C. Parkes. Strategic voting behavior in doodle polls. In *Proceeding of the 18th ACM Conference on Computer Supported Collaborative Work & Social Computing*, pages 464–472. ACM, 2015.