UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Individual Belief Revision Dynamics in a Group Context

Permalink
https://escholarship.org/uc/item/2mg8w82j

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Volzhanin, Igor
Hahn, Ulrike
Jonsson, Martin L
et al.

Publication Date
2015

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Individual Belief Revision Dynamics in a Group Context

Igor Volzhanin (ivolzh01@mail.bbk.ac.uk) and Ulrike Hahn (u.hahn@bbk.ac.uk)
Department of Psychological Sciences, Malet Street
Birkbeck, University of London
London, WC1E 7HX UK

Martin L. Jönsson (martin.jonsson@fil.lu.se) and Erik J. Olsson (erik_j.olsson@fil.lu.se)
Department of Philosophy, Lund University
Lund, 22100 Sweden

Abstract

Our beliefs about the world are generally not formed in isolation: the inherently social nature of human beings means that much of what we believe to know is based, at least in part, on information gained from others. Consequently, human knowledge and its acquisition cannot be fully understood by considering individuals alone. In this paper, we examine the belief dynamics in a group of networked participants engaged in a simple, factual estimation task. Specifically, we examine the extent to which participants revise their own judgments in light of others' responses, and compare formal models of that process.

Keywords: belief revision; social networks; feedback; judgment; advice; social epistemology

Introduction

Social psychology has a long standing interest in group performance and its relationship to the competence of the group's individual members (for reviews see Lorge, 1958; Hill, 1982; Gigone & Hastie,1997). This line of research has received renewed relevance in light of recent developments in network science, which have showed, both through the analysis of real world data and through simulation, how individual behaviour is shaped by the structure of our social networks (see e.g., Jackson 2010). In many contexts, knowing who someone knows is the single best predictor of what they are likely to do (see e.g., Pentland 2014). In keeping with this, philosophers concerned with the nature of knowledge have become increasingly interested in social epistemology (see e.g., Goldman 1999).

It is clear that our social networks influence our behaviour and beliefs, and that individual cognition cannot be fully understood without considering this social dimension. At the same time, a full understanding of how social networks influence individuals cannot proceed without understanding how people respond, at the individual level, to information and cues provided by others. For example, recent simulations suggest the importance of network structure for contagion and diffusion (Kretzschmar & Morris, 1996; Watts, 1999; Lazer & Friedman, 2007) (see Jackson 2010 for an introduction) including such processes as information dissemination (e.g., Doer, Fouz & Friederich, 2012). However, such simulations rest on assumptions about the responses of individual agents. Unless these assumptions match, at least crudely, those of actual people, the insights these models provide remain necessarily limited. Yet, in the cognitive psychological literature on judgment there is remarkably little empirical work on the procedures people use to revise their beliefs in light of information from others in a group (network) setting.

A notable exception are the studies by Yaniv and colleagues (Yaniv, 2004b; Yaniv & Milyavsky, 2007); see also Yaniv 2004a for a review) in the context of the literature on advice. In these studies, participants were given general knowledge questions such as "in what year was the Suez Cana first opened for use?" (Yaniv & Milyavsky, 2007). Participants provided an initial "best estimate" and then received 'advice' from several advisors (e.g., "the best estimate of advisor #33 was 1905"). Participants were then asked to provide a final "best estimate". The main finding is that participants overweight their own opinion relative to that of these (unknown) others: when participants revise their estimates they weight their own answer more strongly than they do the answers of others. This is consistent with results from other advice paradigms such as cue-learning (Harvey & Fischer, 1997), or forecasting (Lim & O'Connor, 1995). In contrast to these other studies, however, Yaniv and Milyavsky also examined the effects of receiving multiple pieces of evidence, each from a different agent. In their (2007) study, participants received an estimate from either 2, 4 or 8 advisors (in actual fact these advisor estimates were drawn randomly from a pool of initial estimates provided by participants in an earlier study). Participants' accuracy improved in all conditions as a result of incorporating such advice, but the benefits of additional estimates seemed to decrease with number. Yaniv and Milyavsky (2007) also examined a range of possible models of participant strategy in their study, finding evidence for discounting of opinions that were too distant from the initial guess. In general, participants seemed sensitive to both their own degree of knowledge, and to how far other opinions were from their own.

While Yaniv and Miyavsky (2007) make an important start in seeking to pin down, on a process level, individual's belief revision in light of information from others, much work remains to be done. For one, the scarcity of studies of this kind makes a replication of interest in and of itself. However, it would also seem desirable to extend the paradigm in a number of other ways. For one, the 'advisors' in Yaniv and Milyavsky's experiments exist from the perspective of participants simply as a minimal verbal label ('advisor #33'). It is unclear to what extent participants consider these advisors to be real people, and what intentions and properties they might attribute. This seems particularly important because one plausible reason for the greater weight placed on participants' own judgments could lie in considerations of source reliability. Participants know a fair bit about themselves and nothing about these other sources (including whether they even exist, other than as an experimental manipulation), and information received from less reliable sources normatively *should* (and does) have less impact on beliefs (see also, Bovens & Hartmann, 2002; Bovens & Hartmann, 2003; Hahn, Harris & Corner, 2009). It would therefore be interesting to conduct such a study in a context where the advisors are clearly other human beings, genuinely engaged in the task at hand. At the same time, Yaniv and Milyavsky examined only one round of advice and subsequent revision, but many social contexts involve repeated exchanges, and hence dynamic interactions whereby our opinions change the beliefs and opinions of others and these influence us in return. We consequently sought to examine behaviour in a general knowledge estimation task, involving multiple, repeated rounds of information exchange between real people.

Belief Revision

An experimental investigation along these lines needs a design in which participants can see that advice is coming from others, yet that is as experimentally controlled as Yaniv and Milyavsky's study. In particular, the experimental context should not introduce a wealth of other factors that might impact the perceived reliability of these information sources. We consequently made use of an experimental context in which a group of participants is present in a room at the same time, but interact only with a computer terminal in front of them. After providing their initial answers, they see the answers of some of the other participants on screen. Participants do not, however, know which answers belong to which person present in the room. Participants then have the opportunity to revise their answers, and this procedure is repeated over several rounds.

The data analysed in this paper come from such a study run at Lund University, Sweden. Its primary aim was to examine the impact of network topology (structure) on the accuracy of participants' beliefs, both individually and collectively. The results of this are presented elsewhere (Jönsson, Hahn, & Olsson, in press). Our interest, here, is in trying to understand the algorithms by which individual participants revise their beliefs. The original study involved gathering significant numbers of trial by trial changes in participants' answers, which allows for detailed analysis of participant behaviour. In the research reported here, we are interested in what strategies people used to revise their behaviour and in testing specific models of that behaviour.

For the purposes of this paper we focus on individual revision statistics in only one of the conditions in the original study. In this condition, participants see the responses of only a subset of the other participants within the group. The information channels that this gives rise to have a small world network structure (Watts & Strogatz, 1998). Small world networks are of interest in this context because many real-world social networks have a small world structure (see also, Watts 1999). This network structure is characterised by short average path lengths between nodes in the network and a higher degree of clustering than seen in the random graph model of Erdos and Rényi (1959). It is a consequence of the partial connectivity in such a network that feedback coming from others is likely to continue to change over more consecutive rounds, regardless of how responses are actually incorporated. By contrast, a complete network – where everyone sees the responses of all other members of the group – would lead to global convergence in a single step if individuals were to adopt the mean of all judgments as their revised answer. The more complex dynamics of small-world networks make them particularly suitable to understanding opinion revision.

Method

Participants 38 undergraduate students (15 male and 23 female) at Lund University took part in the study. They were paid a flat reward for participation (100 SEK), as well as a performance bonus (300 SEK) to the person in each group with the most accurate answers.

Materials & Procedure Participants signed up for one of four sessions, resulting in groups of 9, 9, 7 and 13 participants respectively. The testing conditions, procedure, materials and instructions were the same across the four groups.

During the experiment, each member of a group was seated at a computer and was given two sheets of paper with instructions. When everyone in a group stated that they had understood the instructions, a NetLogo-based program was used to send out questions to all participants. There was an initial warm-up question, followed by ten questions. Each question was repeated eight times over the course of eight consecutive rounds. During the first round each participant answered independently. In the subsequent seven rounds, participants, without prompt, received information about what those they were connected to had answered on the previous round and were asked to revise their answer.

For each question, a small world network was randomly generated (see figure 1). Participants could see the answers of the participants corresponding to the nodes they were immediately connected to. The structure of the network remained the same for the duration of the question. A new network, with new connections was generated for each question. Due to the random nature of network generation, each participant saw at least two, and no more than five answers.

Questions were drawn from a set of 21 questions derived from reports by Statistics Sweden ('Statistiska Centralbyrn') and included questions on Swedish demographics, agriculture and geography. Except for the warm-up question, questions were presented in a random order. All questions asked participants to provide a percentage. Example questions include: "What percent of the Swedes are 15-24 years?" and "What percent of Sweden is covered by agricultural land?"

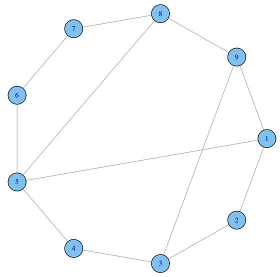

Figure 1: Sample small-world network

All questions thus share a common scale. All groups saw the same warm up question; Groups 1 and 2 saw the same ten questions and Groups 3 and 4 both saw the other ten.

Results

Four scores were eliminated as likely errors before data-analysis. Three were zero-answers that likely resulted from the participant accidentally clicking submit before choosing an estimate (which was done on a sliding bar next to the submit-button); the fourth was a very large number in a sequence of identical low numbers which was also likely to be due to a mis-click.

Rounds Rounds represent discrete time periods that formed the basis for our analysis. Participants had to enter their first answer independently, but were shown other answers in the subsequent rounds. Therefore, belief revision as a result of increased information could be observed in rounds 2 to 8. During the seven rounds of change, participants had an opportunity to enter revised answers, observe others revise their answers and so on.

We used two measures to determine the magnitude of change by each participant: absolute change and percent change. Absolute change refers to by how much each participant changed his or her answer, while percent change refers to the percentage of the change in the subsequent round compared to the previous answer.

Most changes tended to occur in the first round, dropping off sharply and stabilising in the later rounds. As Figure 2 demonstrates, some 35 percent of all change occurred in the first round. This drops off to just under 20 percent in the second round and remains at 10 percent for the later rounds.

This held true for three of the four groups. With respect to the magnitude of change, there were two notable outliers. In Group 4, player 3, revised their answer by 4700 percent in round five, from 3 to 96 (with the correct answer being 96). In Group 3, player 1, changed their answer by 2010 percent, going from 3 to 63 in round two (with correct answer being 55). These two instances are the only changes of this magnitude across all rounds and players. Moreover, these players did not exhibit similar behaviour on other questions. We did not exclude these changes from the overall dataset; however, including them in the graphs significantly distorts the overall picture.

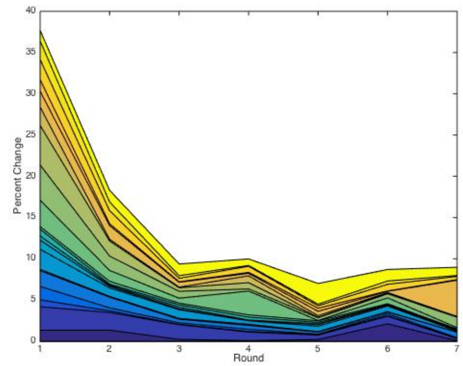

Figure 2: Percentage of Answer Changes by Round (Groups 1-3). Different lines represent different questions.

Percentage Change When we looked at the percentage of change in the answers, we found great variability both for individuals and questions. As an example, Figure 3 breaks down the overall percentage change in answers across rounds for Group 1. Some participants in this group changed their answers by almost 160 percent from their initial value, however, many also did not change their answers at all. The mean value of change was around 20 percent for this group. Figure 4 shows that these same participants also respond very differently to different questions.

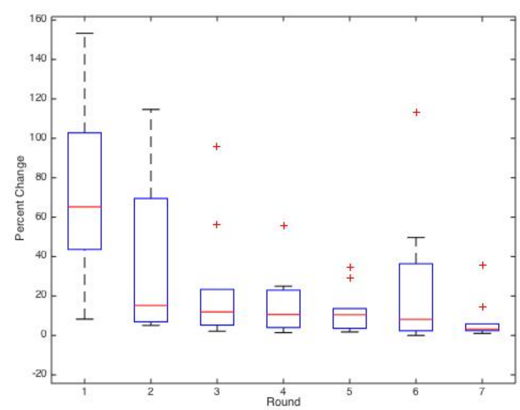

Figure 3: Percent Change of Answers by Round (Group 1)

We also looked at the absolute magnitude of change. In this analysis we added individual changes (and non-changes) for all participants. The histogram in Figure 5 shows that the most prevalent behaviour was not to change the answer at all. In the turns where changes were made, it was mostly by 1 or 2 points. This was true across all groups. As shown in Table 1, the mean absolute change was between 1.1 and 2.3, depending on group.

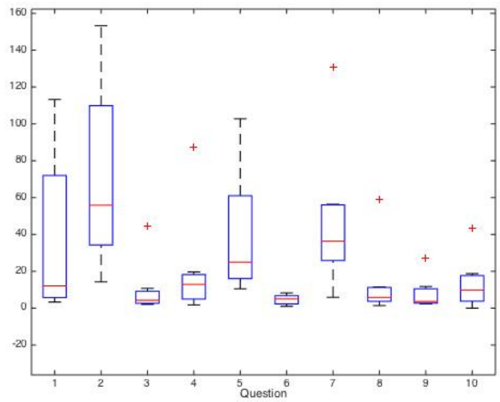

Figure 4: Percent Change of Answers by Question (Group 1)

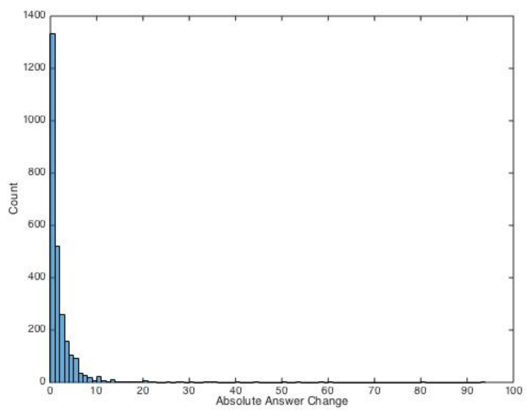

Figure 5: Magnitude of Change Count

Table 1: Mean Absolute Change for all Groups

Group	Mean Change
Group 1	1.1 (SD 2.8)
Group 2	2.3 (SD 3.5)
Group 3	2.2 (SD 2.6)
Group 4	2.1 (SD 2.8)

Discussion

Predictive Models Several models of opinion revision have been proposed. These models typically focus on the individual adopting some combination of the mean, or median values derived from the group. We next describe each of the models we examined in turn. The *weighted average* model predicts that an individual will adopt the group mean, but, in calculating that mean, will weight their own answer more heavily. In our model we set the weight at two: a participant would ‘count’ her own answer twice, before averaging it with the others. In the *split the difference* model, an individual is assumed to take the mean of the others’ answers and then average that with her own answer. The *median* model predicts that the participant will simply adopt the median value of the available answers (including their own). The *no change* model simply predicts that the answer in the next round will be exactly the same as in the previous round. The remaining two models are proposals by Yaniv and Milyavsky (2007) designed to take into account the fact that participants in their study seemed to be sensitive to both the degree of distance of others’ opinions from their own, and, potentially, to the variability within the groups’ judgments. The *egocentric trim* model seeks to capture that participants will, ”weigh the opinions that are close to their own, while ignoring those that are distant from their own prior opinion.” (Yaniv & Milyavsky, 2007, p. 105) In this particular model, an individual will dismiss the value most distant from her own, and adopt the mean value of all remaining answers (including her own). The *consensus trim* model is similar, but here the individual will first take the group mean and then discount the answer most distant from that mean. They will then take another group mean and adopt that value as their own (Yaniv & Milyavsky, 2007). Large distances from other opinions (whether one’s own, or the group mean) may, intuitively be taken to reflect information about source reliability, with ’outliers’ conjectured to likely be less accurate. At the same time, variability within the group may be taken to reflect group confidence. Both these dimensions thus seem worth closer examination.

Yaniv and Milyavsky (2007) in their study tested similar rules – except they tested a straight unweighted mean – instead of our two weighted average rules (weighted average and ’split the difference’); an unweighted mean would fare even worse than the ones we examined, given the extent to which people remained close to their initial opinions. The rules that were the best predictors of behaviour in Yaniv and Milyavsky’s study were the egocentric trim and the median (the latter being the rule that would have also brought participants in their study the greatest gains in accuracy). Moreover, in their study, median, mean, consensus and egocentric trim were more accurate than a no change model.

Table 2 summarises the results for all models on our data. The value shown in a given cell is the mean absolute error per turn (i.e., the mean deviation between predicted value and actual responses for an individual participant on a given round of one question). Higher numbers indicate greater deviation between predicted and observed behaviour.

We first examined the models’ performance on the initial round of change, where comparison is most direct with Yaniv and Milyavsky’s study (which gave participants only one set of advice and thus sought only one revision). In marked contrast to Yaniv and Milyavsky’s results, the ’no change’ model has the lowest predictive error for our data. This suggests a very notable difference in how participants responded in our study. Quite possibly, task demands in Yaniv and Milyavsky’s (2007) study, where the pieces of advice were experimenter provided, were somewhat higher. This emphasises the need to examine social belief dynamics in a broader range of ex-

Table 2: Model Performance

	First Round of Revision Only	Across All Rounds				
Model	Average Across Groups	Group 1	Group 2	Group 3	Group 4	Average Across Groups
Weighted Average	7.31	6.39	7.79	8.74	7.61	7.63
Split the Difference	5.69	5.62	7.22	8.11	6.92	6.97
Median	8.23	5.21	7.77	8.66	6.04	6.92
No Change	5.11	4.03	6.41	6.92	6.06	5.85
Egocentric Trim	6.23	4.41	6.86	8.16	6.23	6.41
Consensus Trim	9.10	5.64	8.13	10.68	6.61	7.76

perimental paradigms. Secondly, the fact that the 'no change' model is the best predictor suggests immediately that none of the models are terribly good. There is significant, and systematic, change in participants' responses, yet all the models seeking to capture this change do less well.

Where our results do fit with Yaniv and Milyavsky's findings is in the rank order of the other three models they test (median, egocentric trim, consensus trim). The consensus trim model performs worst, with the median second, and Yaniv and Milyavsky's (2007) egocentric trim model is the best in both their and our study (n.b. the relationship between median and egocentric trim in their study varies as a function of number of pieces of advice, so we considered the average performance across their conditions, which is also appropriate because participants in our small world network vary in the number of others they are connected to.) Finally, our two averaging models place fourth and fifth.

How then do these models fare in predicting repeated rounds of revision in a dynamically changing environment, where –due to the partial connectivity of the network– information only gradually propagates through the network? Again, the 'no change' model is the best predictor, followed somewhat more closely by the egocentric trim model. Median and split the different are now virtually tied for third. The consensus trim, again, comes last suggesting that it fails to capture participants' approach to opinion variability in a meaningful way.

Where do the failures of the models lie? First, all models (with the obvious exception of the no change model) over-predict change for the first round of revision, that is, the transition from participants initial answer to their second answer. In other words, despite the fact that most change in participants' responses occurs in the first round of revision, participants still change less than the models suggest they should. This can be seen by comparing Figure 6 which displays round on round change for Group 1 participants with Figures 7, 8, and 9. Even the best of the models, Yaniv and Milyavsky's egocentric trim model (Fig. 7), predicts noticeably more change in this round than actually occurs. However, the same models then *under-predict* change on the second round of revision. It appears that repeated feedback encourages participants to take comparatively greater note of other's opinions on this second revision round. This is particularly clear in the comparison with the weighted average model, Fig. 9, which predicts a sharp, monotonic, decrease in round-on-round change.

In other words, there is some suggestion from these comparisons that 'weights' of other's opinions are dynamic, as opposed to unchanging, across the subsequent rounds. Seeking to probe the nature of such dynamic changes further seems imperative given that the most common models of belief and opinion dynamics assume constant weights.

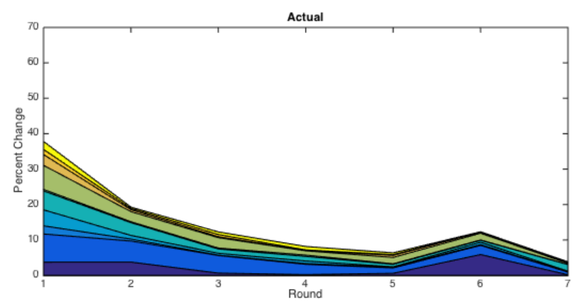

Figure 6: Actual Percentage Change by Round (Group 1)

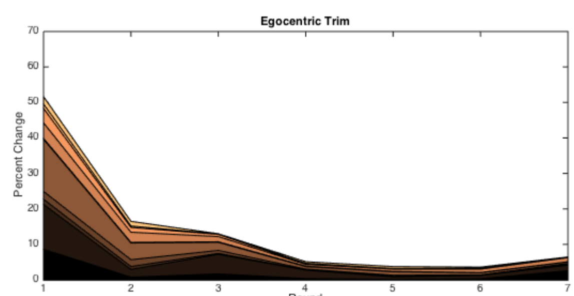

Figure 7: Predicted Percentage Change by Round for Egocentric Trim Model (Group 1)

Conclusion

Although group behaviour has been much studied, individual behaviour within a group remains poorly understood. The fact that the 'no change' model is the best predictor of participants' revision, when there are clearly considerable amounts

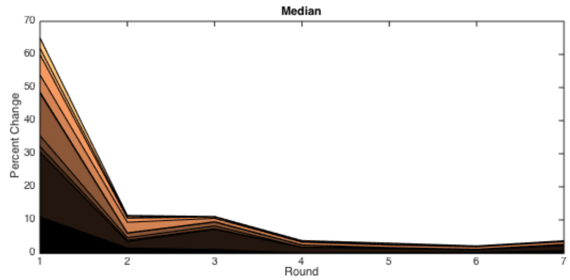

Figure 8: Predicted Percentage Change by Round for Median Model (Group 1)

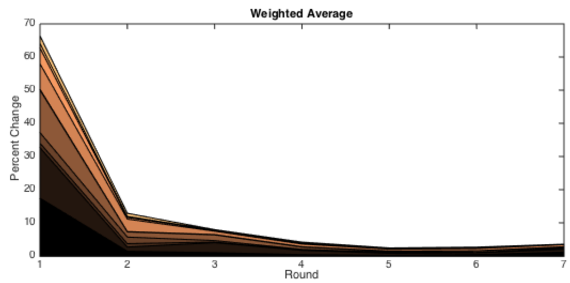

Figure 9: Predicted Percentage Change by Round for Weighted Average Model (Group 1)

of change in participants' estimates, suggests that all the models examined here, and with that the main models suggested by the experimental literature, are still well off the mark. At the same time, there is some convergence with past results in that the best of the models we considered is Yaniv and Milyavsky's (2007) 'egocentric trim' model, which was also the (overall) best performing model in their study. This suggests that this model provides a good starting point for the development of better models. At the same time, it remains intuitive that people should be sensitive not just to the distance of other's opinions to their own, but also to the variability among opinions. The consensus model of Yaniv and Milyavsky (2007) was the only model tested that incorporated sensitivity to variability, and it was by far the worst performing. This suggests that further exploration of people's sensitivity to variability, both experimentally and through modelling, is an important avenue for further research.

Acknowledgments

The research reported was partially supported by two Swedish Research Council grants: the Hesselgren Professorship (U. Hahn), and the framework project "Knowledge in a Digital World" (E. J. Olsson, PI; M. Jönsson, I. Volzhanin).

References

Bovens, L., & Hartmann, S. (2002, March). Bayesian Networks and the Problem of Unreliable Instruments. *Philosophy of Science*, *69*(1), 29–72.

Bovens, L., & Hartmann, S. (2003). Bayesian epistemology.

Doer, B., Fouz, M., & Friedrich, T. (2012, June). Why rumors spread so quickly in social networks. *Communications of the ACM*, *55*(6), 70.

Erdos, P., & Rényi, I. (1959). On Random Graphs, I. *Publicationes mathematicae*, *6*, 290–297.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*(1), 149.

Goldman, A. I. (1999). *Knowledge in a social world*. Oxford University Press.

Hahn, U., Harris, A., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*, 337–367.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133.

Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin*, *91*(3), 517.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Jönsson, M. L., Hahn, U., & Olsson, E. J. (in press). The Kind of Group YouWant to Belong to: Effects of Group Structure on Group Accuracy. *Cognition*.

Kretzschmar, M., & Morris, M. (1996, April). Measures of concurrency in networks and the spread of infectious disease. *Mathematical biosciences*, *133*(2), 165–195.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667–694.

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*(3), 149–168.

Lorge, I., & Brenner, M. (1958). A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance: 1920-1957. *Psychological Bulletin*, *55*, 337–372.

Pentland, A. (2014). *Social Physics: how good ideas spread – the lessons from a new science*. Scribe Publications.

Watts, D. J. (1999). Networks, Dynamics, and the Small World Phenomenon. *American Journal of Sociology*, *105*(2), 493–527.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature*, *393*(6684), 440–442.

Yaniv, I. (2004a). The Benefit of Additional Opinions. *Current Directions in Psychological Science*, *13*, 75–78.

Yaniv, I. (2004b). *Receiving other people's advice: Influence and benefit* (Vol. 93). Organizational Behavior and Human Decision Processes.

Yaniv, I., & Milyavsky, M. (2007, May). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104–120.