The wisdom of networked crowds

Mark C. Wilson & Valery Pavlov
University of Auckland

CMSS Summer Workshop, 20 February 2016

Main research questions

- How do agents update their beliefs in the context of a social network?

Main research questions

- How do agents update their beliefs in the context of a social network?
- Do they change directly from "believing X" to "believing not-X" directly, or do they first become undecided?

Main research questions

- How do agents update their beliefs in the context of a social network?
- Do they change directly from “believing X” to “believing not-X” directly, or do they first become undecided?
- Is the change of belief better modelled as a simple contagion or a complex contagion?

Main research questions

- How do agents update their beliefs in the context of a social network?
- Do they change directly from "believing X" to "believing not-X" directly, or do they first become undecided?
- Is the change of belief better modelled as a simple contagion or a complex contagion?
- Are threshold models appropriate?

Main research questions

- How do agents update their beliefs in the context of a social network?
- Do they change directly from “believing X” to “believing not-X” directly, or do they first become undecided?
- Is the change of belief better modelled as a simple contagion or a complex contagion?
- Are threshold models appropriate?
- Which classes of diffusion models definitely don’t work, and which are promising?

Wisdom of crowds

- Has been studied since 1907 when Galton analysed the estimates of 787 people of the weight of an ox.

Wisdom of crowds

- Has been studied since 1907 when Galton analysed the estimates of 787 people of the weight of an ox.
- Key finding: the mean of estimates is usually more accurate than any single estimate: the crowd is wiser than any of its members, including "experts".

Wisdom of crowds

- Has been studied since 1907 when Galton analysed the estimates of 787 people of the weight of an ox.
- Key finding: the mean of estimates is usually more accurate than any single estimate: the crowd is wiser than any of its members, including "experts".
- Theoretical justification for this comes from the Law of Large Numbers, with appropriate assumptions on distribution of the errors made by individuals.

Wisdom of crowds

- Has been studied since 1907 when Galton analysed the estimates of 787 people of the weight of an ox.
- Key finding: the mean of estimates is usually more accurate than any single estimate: the crowd is wiser than any of its members, including "experts".
- Theoretical justification for this comes from the Law of Large Numbers, with appropriate assumptions on distribution of the errors made by individuals.
- Independence of estimates (or even negative correlation) gives good results, but positively correlated estimates can give bad ones, in theory.

Wisdom of crowds

- Has been studied since 1907 when Galton analysed the estimates of 787 people of the weight of an ox.
- Key finding: the mean of estimates is usually more accurate than any single estimate: the crowd is wiser than any of its members, including "experts".
- Theoretical justification for this comes from the Law of Large Numbers, with appropriate assumptions on distribution of the errors made by individuals.
- Independence of estimates (or even negative correlation) gives good results, but positively correlated estimates can give bad ones, in theory.
- When estimates can be revised and information about others' estimates is available, herding can occur.

Decision-making in groups

- There is a large body of work in social psychology and experimental economics involving decision-making in groups.

Decision-making in groups

- There is a large body of work in social psychology and experimental economics involving decision-making in groups.
- Sometimes there are strategic incentives for conformity rather than correctness. We focus on the case where individual performance is incentivized and answers are objective.

Decision-making in groups

- There is a large body of work in social psychology and experimental economics involving decision-making in groups.
- Sometimes there are strategic incentives for conformity rather than correctness. We focus on the case where individual performance is incentivized and answers are objective.
- Group performance depends greatly on protocols used (for example, discussion can often hinder rather than help). The Delphi technique is a generally successful method for combining estimates allowing for iterative guesses based on feedback from the group, conveyed via a central controller.

Our problem

- We focus on the iterative distributed jury situation (our term):

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized
  - multiple choice answers

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized
  - multiple choice answers
  - discussion is not possible or allowed

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized
  - multiple choice answers
  - discussion is not possible or allowed
  - participants iteratively revise estimates based on polls

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized
  - multiple choice answers
  - discussion is not possible or allowed
  - participants iteratively revise estimates based on polls
  - information flow is controlled by a network

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized
  - multiple choice answers
  - discussion is not possible or allowed
  - participants iteratively revise estimates based on polls
  - information flow is controlled by a network
- We found little literature on this situation. Without the iteration and network, Condorcet's Jury Theorem and its descendants show that crowds are often wiser than their experts.

Our problem

- We focus on the iterative distributed jury situation (our term):
  - truth is incentivized
  - multiple choice answers
  - discussion is not possible or allowed
  - participants iteratively revise estimates based on polls
  - information flow is controlled by a network
- We found little literature on this situation. Without the iteration and network, Condorcet's Jury Theorem and its descendants show that crowds are often wiser than their experts.
- We seek to investigate this experimentally with a view to developing more realistic models that we can analyse. In particular, we want to model undecidedness.

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.
- Allowed third option ("I am not sure") and incentivized choosing this over randomly guessing one of the two provided answers.

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.
- Allowed third option ("I am not sure") and incentivized choosing this over randomly guessing one of the two provided answers.
- This allows us to see more detail on how beliefs are updated.

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.
- Allowed third option ("I am not sure") and incentivized choosing this over randomly guessing one of the two provided answers.
- This allows us to see more detail on how beliefs are updated.
- It also allows us to focus on three subgroups:

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.
- Allowed third option ("I am not sure") and incentivized choosing this over randomly guessing one of the two provided answers.
- This allows us to see more detail on how beliefs are updated.
- It also allows us to focus on three subgroups:
  - those who know they know;

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.
- Allowed third option ("I am not sure") and incentivized choosing this over randomly guessing one of the two provided answers.
- This allows us to see more detail on how beliefs are updated.
- It also allows us to focus on three subgroups:
  - those who know they know;
  - those who know they don't know;

Our experiment

- A standard behavioral lab experiment with randomly selected undergraduate participants.
- Allowed third option ("I am not sure") and incentivized choosing this over randomly guessing one of the two provided answers.
- This allows us to see more detail on how beliefs are updated.
- It also allows us to focus on three subgroups:
  - those who know they know;
  - those who know they don't know;
  - those who don't know they don't know.

What the subjects saw

Question

1 out of 5

Remaining time to make your choice 14

This is iteration 2.
The percentages of your feeds who chose answers 1), 2) or 3) were 0%, 0%, 100%, respectively. Your previous answer to this question was 2. (Note: "0" means you did not answer).

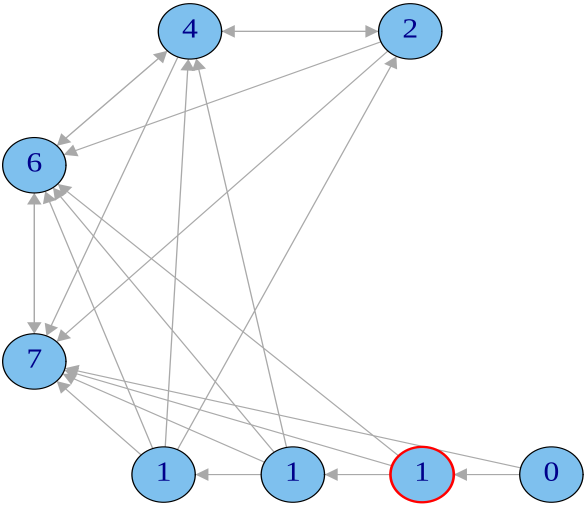
Question

If it takes 5 machines 5 minutes to make 5 widgets, how long will it take 100 machines to make 100 widgets?

Options

1) At least 50 minutes
2) Less than 50 minutes
3) I am not sure

Novel topology creating heterogeneity in feed information



Types of questions

- We chose a variety of questions. In particular:

Types of questions

- We chose a variety of questions. In particular:
  - Two were purely logical ("intellective")in nature (one from Frederick's Cognitive Reflection Test and a variant of the Wason task).

Types of questions

- We chose a variety of questions. In particular:
  - Two were purely logical ("intellective")in nature (one from Frederick's Cognitive Reflection Test and a variant of the Wason task).
  - Two required knowledge of facts about the world (movie actors, a common misconception about geography).

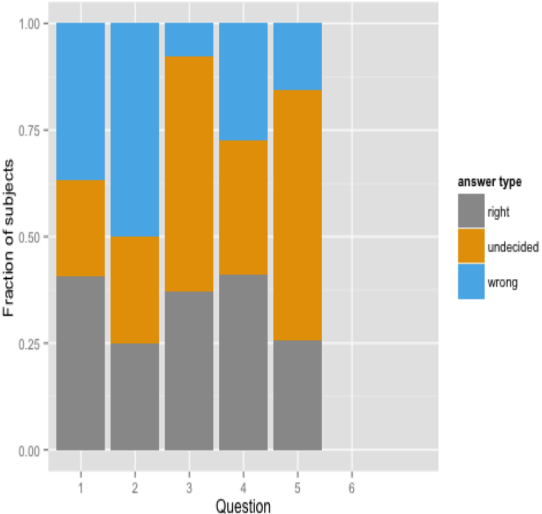
Types of questions

- We chose a variety of questions. In particular:
  - Two were purely logical ("intellective")in nature (one from Frederick's Cognitive Reflection Test and a variant of the Wason task).
  - Two required knowledge of facts about the world (movie actors, a common misconception about geography).
  - One was essentially impossible without being given the correct answer (which we gave to a few subjects).

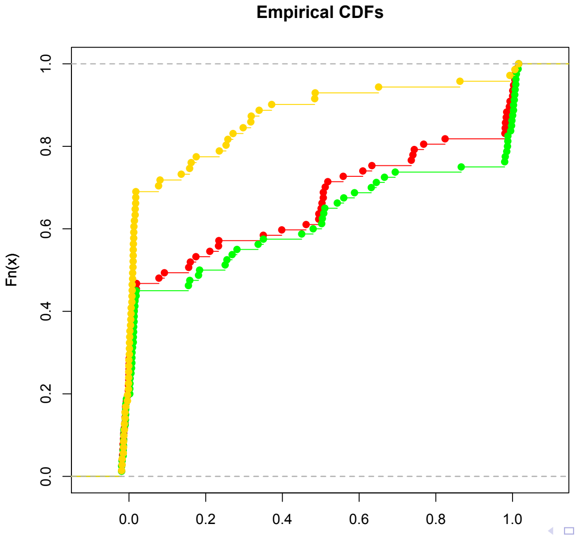
Types of questions

- We chose a variety of questions. In particular:
  - Two were purely logical ("intellective")in nature (one from Frederick's Cognitive Reflection Test and a variant of the Wason task).
  - Two required knowledge of facts about the world (movie actors, a common misconception about geography).
  - One was essentially impossible without being given the correct answer (which we gave to a few subjects).
- Let C, U, I be the fraction of correct, undecided, incorrect answers given at the first iteration, so $1 = C + U + I$. A key property of questions seems to be trickiness, which we define as I. This is distinct from difficulty, which we define as $U + I$.

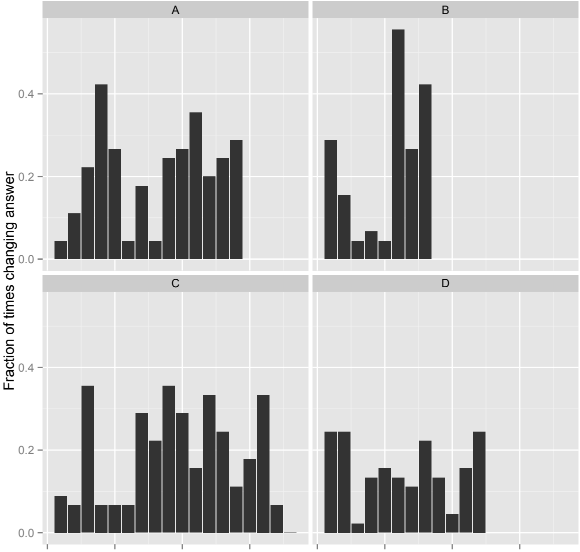
Difficulty and trickiness of questions



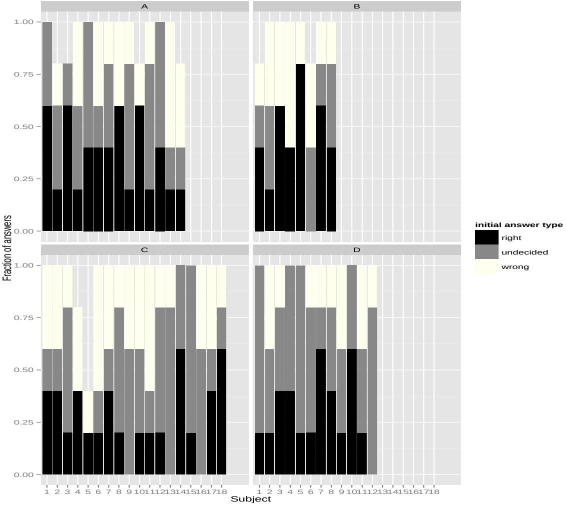
Strong asymmetry between "I am not sure" and others



Participants changed their answer rather often



About 80% were tricked on some question



Finding — importance of question type

- There are large differences in answer patterns between questions.

Finding — importance of question type

- There are large differences in answer patterns between questions.
- The higher the trickiness of the question, the worse the social learning.

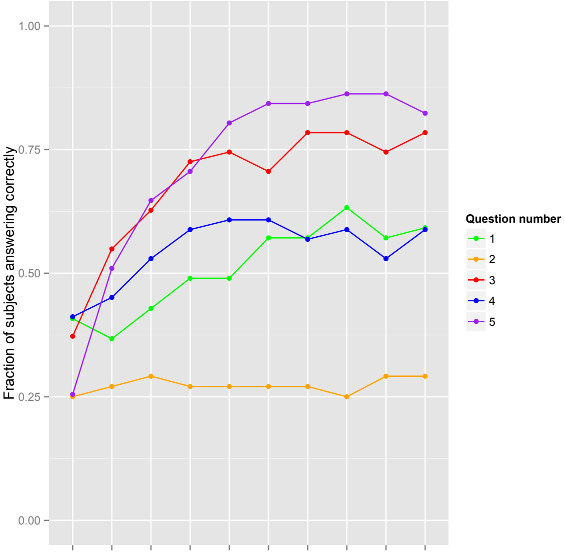
Finding — importance of question type

- There are large differences in answer patterns between questions.
- The higher the trickiness of the question, the worse the social learning.
- Difficulty is not as important as trickiness. Learning can occur well on difficult questions.

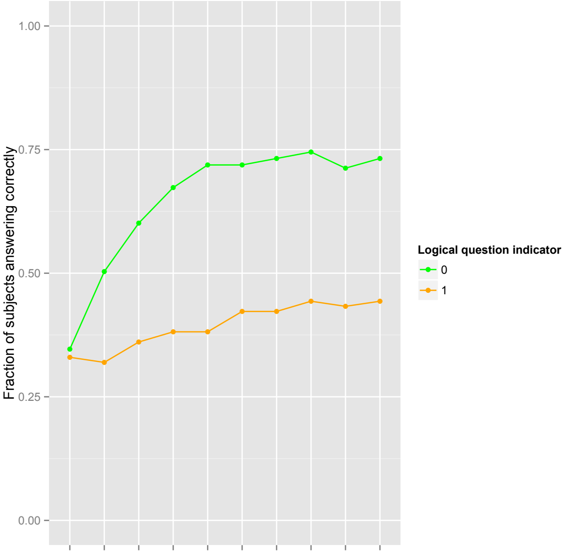
Finding — importance of question type

- There are large differences in answer patterns between questions.
- The higher the trickiness of the question, the worse the social learning.
- Difficulty is not as important as trickiness. Learning can occur well on difficult questions.
- Logical questions elicit more changes but worse social learning.

Aggregate learning by question



Aggregate learning by question type



Finding — self-awareness promotes learning

- There is a clear difference in learning between those who are correct at the first iteration, those who answer wrongly, and those who admit to being undecided.

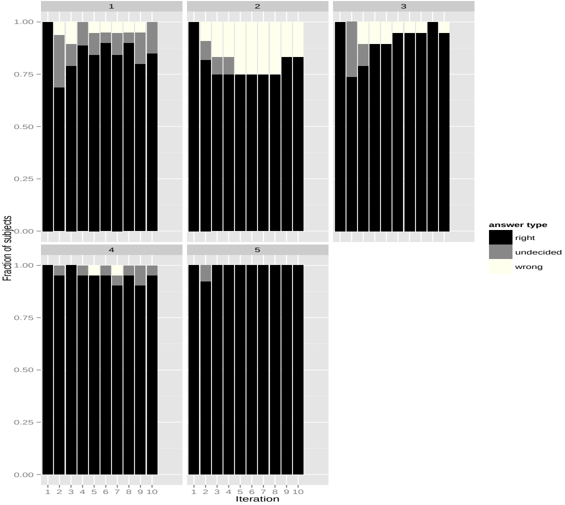
Finding — self-awareness promotes learning

- There is a clear difference in learning between those who are correct at the first iteration, those who answer wrongly, and those who admit to being undecided.
- This effect even occurred (to a lesser extent) in the complete topology, where everyone has the same information.

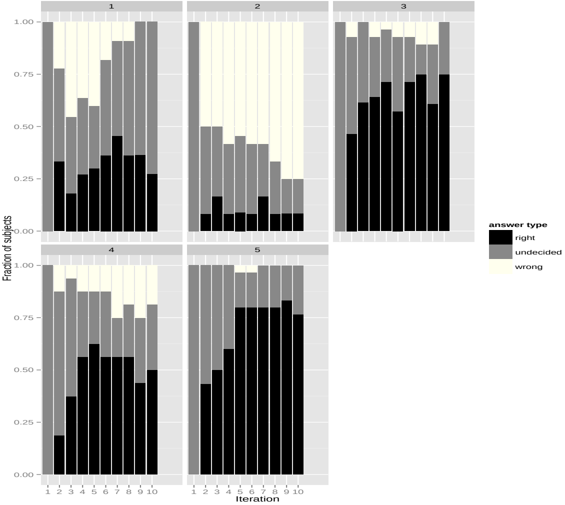
Finding — self-awareness promotes learning

- There is a clear difference in learning between those who are correct at the first iteration, those who answer wrongly, and those who admit to being undecided.
- This effect even occurred (to a lesser extent) in the complete topology, where everyone has the same information.
- It does not seem that the members of these groups are the same for each question, but more work is needed.

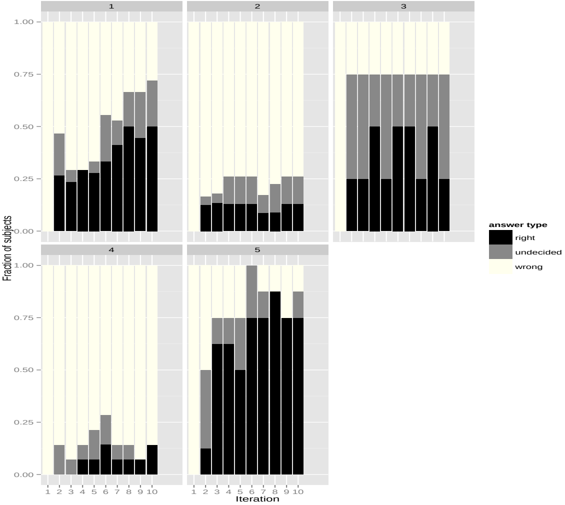
Those who (know that they) know



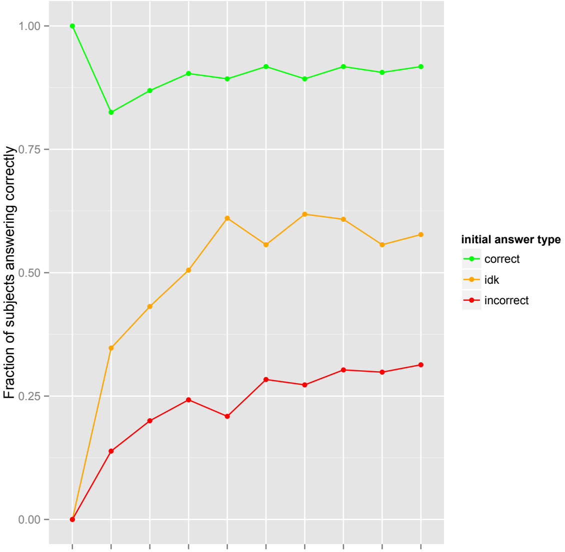
Those who know that they don't know



Those who don't know that they don't know



Learning according to first answer



Finding — effect of topology

- For the high-trickiness Question 2, there is a huge difference in dynamics between the two network topologies we used, otherwise moderate difference.

Finding — effect of topology

- For the high-trickiness Question 2, there is a huge difference in dynamics between the two network topologies we used, otherwise moderate difference.
- In the "spiral" topology, convergence is slow or nonexistent. Convergence is much faster in the complete topology.

Finding — effect of topology

- For the high-trickiness Question 2, there is a huge difference in dynamics between the two network topologies we used, otherwise moderate difference.
- In the "spiral" topology, convergence is slow or nonexistent. Convergence is much faster in the complete topology.
- For the less tricky questions, convergence to truth occurs in the complete topology, but in the high-trickiness question, convergence to the wrong answer occurs.

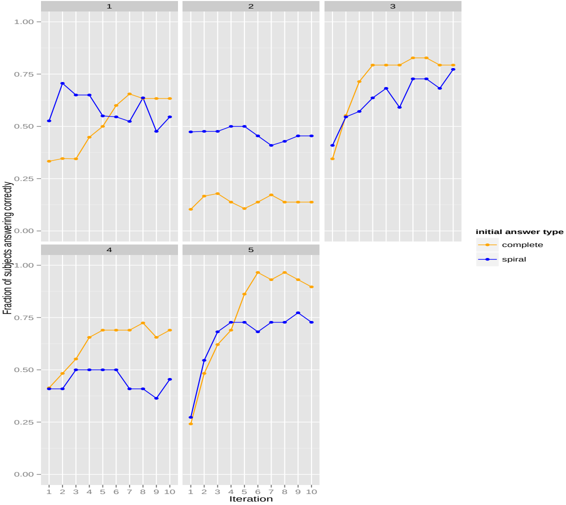
Finding — effect of topology

- For the high-trickiness Question 2, there is a huge difference in dynamics between the two network topologies we used, otherwise moderate difference.
- In the "spiral" topology, convergence is slow or nonexistent. Convergence is much faster in the complete topology.
- For the less tricky questions, convergence to truth occurs in the complete topology, but in the high-trickiness question, convergence to the wrong answer occurs.
- Our results, in quite a different setting, seem to agree with those of Lazer & Friedman (2007) on exploration vs exploitation in optimization problems.

Finding — effect of topology

- For the high-trickiness Question 2, there is a huge difference in dynamics between the two network topologies we used, otherwise moderate difference.
- In the “spiral” topology, convergence is slow or nonexistent. Convergence is much faster in the complete topology.
- For the less tricky questions, convergence to truth occurs in the complete topology, but in the high-trickiness question, convergence to the wrong answer occurs.
- Our results, in quite a different setting, seem to agree with those of Lazer & Friedman (2007) on exploration vs exploitation in optimization problems.
- Topologies promoting easy information-sharing tend to yield worse social learning for hard questions, as too little exploration occurs owing to herding.

Learning by question and topology



Implications for modelling

- Condorcet Jury Theorem-type results where individual correctness probability is replaced by incorrectness.

Implications for modelling

- Condorcet Jury Theorem-type results where individual correctness probability is replaced by incorrectness.
- Models predicting convergence to unanimity are inconsistent with our data.

Implications for modelling

- Condorcet Jury Theorem-type results where individual correctness probability is replaced by incorrectness.
- Models predicting convergence to unanimity are inconsistent with our data.
- Data shows that simple (epidemic-like) contagion typical of information models do not fit the data as well as complex (threshold-type) models of behavioral change.

Implications for modelling

- Condorcet Jury Theorem-type results where individual correctness probability is replaced by incorrectness.
- Models predicting convergence to unanimity are inconsistent with our data.
- Data shows that simple (epidemic-like) contagion typical of information models do not fit the data as well as complex (threshold-type) models of behavioral change.
- Threshold-type models for opinion change won't work directly: any model with zero weight on agent's own opinion is strongly refuted by our data.

Implications for modelling

- Condorcet Jury Theorem-type results where individual correctness probability is replaced by incorrectness.
- Models predicting convergence to unanimity are inconsistent with our data.
- Data shows that simple (epidemic-like) contagion typical of information models do not fit the data as well as complex (threshold-type) models of behavioral change.
- Threshold-type models for opinion change won't work directly: any model with zero weight on agent's own opinion is strongly refuted by our data.
- If p_1, p_2, p_3 are fractions of neighbours with opinions $1, 2$ or "I am not sure", then threshold models use p_1 to predict answer 1 by agent. However we find that $p_1 - p_2$ and p_3 are much better predictors.

Experimental setup I

- Used DECIDE behavioral laboratory at University of Auckland (`http://hfbeltran.wix.com/decide`) run by A. Chaudhuri, Pavlov and H. Beltran.

Experimental setup I

- Used DECIDE behavioral laboratory at University of Auckland (`http://hfbeltran.wix.com/decide`) run by A. Chaudhuri, Pavlov and H. Beltran.
- A standard setup with participants sitting at networked computers, no communication except via the experimenters. Subjects randomly assigned, topology determined by us (one of two types), nodes randomly permuted between questions.

Experimental setup I

- Used DECIDE behavioral laboratory at University of Auckland (`http://hfbeltran.wix.com/decide`) run by A. Chaudhuri, Pavlov and H. Beltran.
- A standard setup with participants sitting at networked computers, no communication except via the experimenters. Subjects randomly assigned, topology determined by us (one of two types), nodes randomly permuted between questions.
- Standard software ORSEE used to recruit from a pool of students, zTree used to perform the experiment.

Experimental setup II

- Several *multiple choice* questions are answered, one at a time — other studies mostly use free-form answers.

Experimental setup II

- Several *multiple choice* questions are answered, one at a time — other studies mostly use free-form answers.
- Each question has 10 iterations, and subjects can change their answer each time if desired. They receive summary feedback on their neighbours' answers after each iteration.

Experimental setup II

- Several *multiple choice* questions are answered, one at a time — other studies mostly use free-form answers.
- Each question has 10 iterations, and subjects can change their answer each time if desired. They receive summary feedback on their neighbours' answers after each iteration.
- The first iteration is allotted 2 minutes, each subsequent one 30 seconds.

Experimental setup II

- Several *multiple choice* questions are answered, one at a time — other studies mostly use free-form answers.
- Each question has 10 iterations, and subjects can change their answer each time if desired. They receive summary feedback on their neighbours' answers after each iteration.
- The first iteration is allotted 2 minutes, each subsequent one 30 seconds.
- There are 3 possible answers: option 1, option 2 and "I am not sure". This we think is novel.

Experimental setup II

- Several *multiple choice* questions are answered, one at a time — other studies mostly use free-form answers.
- Each question has 10 iterations, and subjects can change their answer each time if desired. They receive summary feedback on their neighbours' answers after each iteration.
- The first iteration is allotted 2 minutes, each subsequent one 30 seconds.
- There are 3 possible answers: option 1, option 2 and "I am not sure". This we think is novel.
- Subjects are paid for answering correctly at the last iteration and a randomly chosen other one. They are paid 1 for correct, 0 for wrong, 0.6 for "I am not sure".

Data

- Each data point is an answer to a fixed question at a specific iteration by a specific subject. We also measure time to answer and the summary of answers of the neighbours in the previous iteration.

Data

- Each data point is an answer to a fixed question at a specific iteration by a specific subject. We also measure time to answer and the summary of answers of the neighbours in the previous iteration.
- We have 52 subjects over 4 sessions, so 2600 data points. In only 72 of these was no answer made. We discard those points for most of the analysis, leaving 2528.

Key references

- *How social influence can undermine the wisdom of crowd effect*: Lorenz, Rauhut, Schweitzer, Helbing, PNAS 2011.

Key references

- *How social influence can undermine the wisdom of crowd effect*: Lorenz, Rauhut, Schweitzer, Helbing, PNAS 2011.
- *Analytical reasoning task reveals limits of social learning in networks*: Rahwan, Krasnoshtan, Shariff, Bonnefon. J. R. Soc. Interface 2014.

Key references

- *How social influence can undermine the wisdom of crowd effect*: Lorenz, Rauhut, Schweitzer, Helbing, PNAS 2011.
- *Analytical reasoning task reveals limits of social learning in networks*: Rahwan, Krasnoshtan, Shariff, Bonnefon. J. R. Soc. Interface 2014.
- *The Network Structure of Exploration and Exploitation*: Lazer, Friedman. Administrative Science Quarterly 2007.