

Crowd Wisdom and Deliberative Voting: Does Deliberation Improve Estimation Accuracy?

Bernhard Kittel¹, Jan Lorenz¹, Heiko Rauhut²

¹ Center for Social Science Methodology, Carl von Ossietzky Universität Oldenburg

² Soziologisches Institut, Universität Zürich

Gründungstagung der Sektion Methoden der Politikwissenschaft
der Deutschen Vereinigung für Politische Wissenschaft (DVPW), Hamburg

February 11, 2012

Expert committees have two tasks

- ▶ Expert committees shall
 - ▶ find the **best possible predictions**
 - ▶ deliver a **common opinion**
- ▶ Mixed task of scientific and political **deliberation**
- ▶ Example: **Intergovernmental Panel on Climate Change**
 - ▶ Shall predict sea level 2050 and 2100. Correctness is important:
 - ▶ Building dikes without need would be a waste of money ...
 - ▶ High sea level and no dikes will cause costly disasters ...
 - ▶ Prevention of CO₂-emissions would be a waste of money if sea level rises anyway ...
 - ▶ Described as “bazaar” where future sea level is “negotiated”.¹
 - ▶ Discusses its internal report and decision structures.²

¹ Klimarat feilscht um Daten zum Meeresspiegel-Anstieg, Axel Bojanowski, Spiegel Online, 14.07.2011

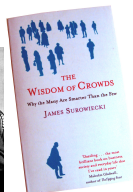
² Experten drängen auf Umbau des Weltklimarats, Axel Bojanowski, Spiegel Online, 30.08.2010

The Wisdom of the Crowd

- Situation:**
1. A question about an unknown fact of metric nature
 2. Estimates from many individuals

- Wisdom-of-Crowd Effect:**
1. An aggregate of all estimates is “surprisingly” close to the true value.
 2. Many estimates are far away from the truth

- ▶ Galton (1907): The median of a weight judging competition outperformed **every individual estimate**.
- ▶ Popularized by Surowiecki 2004



WoC-Effect present: **Ask many independently and aggregate!**

(And the law of large numbers ensures good answers.)

Research Question

- ▶ Do expert committees use the wisdom-of-crowd effect?
- ▶ If yes, **why deliberation and collective decision?**
(It would undermine independence.)

Social influence by exchange of information can undermine the effect³. But exchange can not be avoided in an open society.

- ▶ Can expert committees use the WoC effect **despite** deliberation?
- ▶ Can expert committees even **improve** the outcome beyond the WoC effect?
- ▶ What's the impact of collective decisions?

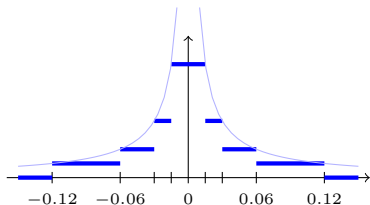
³ J.Lorenz, H.Rauhut, F.Schweitzer, D.Helbing How social influence undermines the wisdom of crowd effect, PNAS 2011

Experimental Design

Operationalize expert committee's task in laboratory experiment:

- ▶ Prediction \leftrightarrow **Factual Questions**. Example:
What percentage of Saudi's oil is used in Saudi Arabia?
- ▶ Expect correct predictions \leftrightarrow **incentivize correct answers**

Deviation	Payment
$\leq 1.5\%$	€4
$\leq 3\%$	€2
$\leq 6\%$	€1
$\leq 12\%$	€0.5
larger	€0



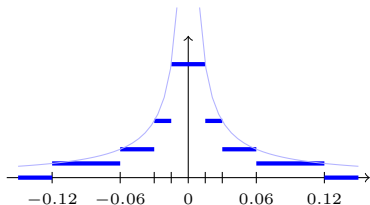
- ▶ Strong incentive to come closer to the truth!

Experimental Design

Operationalize expert committee's task in laboratory experiment:

- ▶ Prediction \leftrightarrow **Factual Questions**. Example:
What percentage of Saudi's oil is used in Saudi Arabia?
 - ▶ **19 %**
- ▶ Expect correct predictions \leftrightarrow **incentivize correct answers**

Deviation	Payment
$\leq 1.5\%$	€4
$\leq 3\%$	€2
$\leq 6\%$	€1
$\leq 12\%$	€0.5
larger	€0



- ▶ Strong incentive to come closer to the truth!

Experimental Design

- ▶ Invited for 6 **sessions** of 24 participants each (one with 21)
- ▶ In each session 12 **rounds** with 12 questions (from 18)
- ▶ **Decision rules**
 1. **Individual rule** ← no group decision, all estimates in report
 2. **Majority rule** ← common estimate desired, even disagreed
 3. **Consensus rule** ← agreement desired/necessary
 - ▶ No group decision (in 2./3.) ← committee fails to deliver report

Chat interface all-to-all, free text

Verbleibende Zeit [sec]: 1:45

Runde 1 von 1. Sie sind **M1** und spielen mit **9 Personen**. Entscheidungs-Regel: **Mehrheit**. Kommunikation: **im Chat**.

Frage:

Kommunikation im Chat (Zum Chatten tippen Sie Ihre Nachricht ein und drücken Sie ENTER)

Die Chatbox verschwindet nach 300 Sekunden! Wenn alle Mitspielenden bereit sind Ihre Entscheidung einzugeben, brauchen Sie nicht die ganze Zeit abzuwarten.

M8: Ich glaub 80
M8: 70
M9: keine Ahnung
M5: hätte jetzt 50% gesagt...

Hmmm...

Anzahl von 9 Personen die bereit sind, eine Entscheidung einzugeben: 2

Ihre Bereitschaft eine Entscheidung einzugeben: Ja Nein

Ich bin bereit eine Entscheidung einzugeben

Entscheidung

Sobald die Chat-Zeit abgelaufen ist oder alle Mitspielenden bereit sind Ihre Entscheidung einzugeben können sie hier Ihre Entscheidung eingeben. Entscheidungsregel: **Mehrheit**

<p>Entscheidungs-Regeln</p> <p>Individual: Die Kommunikation dient nur dem Austausch, Sie treffen eine individuelle Entscheidung. Es muss keine Gruppenentscheidung gefunden werden.</p> <p>Mehrheit: In Ihrer Gruppe muss sich eine Mehrheit bilden, so dass mehr als die Hälfte der Mitglieder die gleiche Entscheidung eingibt. Gelingt das nicht, gibt es keine Auszahlung.</p> <p>Konsens: In Ihrer Gruppe müssen alle Mitglieder die gleiche Entscheidung eingeben. Gelingt das nicht, gibt es keine Auszahlung.</p>	<p>Auszahlungs-Intervalle</p> <p>Auszahlungen richten sich nach dem Abstand der Entscheidung zum richtigen Wert.</p> <table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 2px;">Abstand kleiner oder gleich 1,5</td> <td style="padding: 2px;">-></td> <td style="padding: 2px;">4 Punkte</td> </tr> <tr> <td style="padding: 2px;">Abstand kleiner oder gleich 3</td> <td style="padding: 2px;">-></td> <td style="padding: 2px;">2 Punkte</td> </tr> <tr> <td style="padding: 2px;">Abstand kleiner oder gleich 6</td> <td style="padding: 2px;">-></td> <td style="padding: 2px;">1 Punkte</td> </tr> <tr> <td style="padding: 2px;">Abstand kleiner oder gleich 12</td> <td style="padding: 2px;">-></td> <td style="padding: 2px;">0,5 Punkte</td> </tr> <tr> <td style="padding: 2px;">Abstand größer 12</td> <td style="padding: 2px;">-></td> <td style="padding: 2px;">0 Punkte</td> </tr> </table>	Abstand kleiner oder gleich 1,5	->	4 Punkte	Abstand kleiner oder gleich 3	->	2 Punkte	Abstand kleiner oder gleich 6	->	1 Punkte	Abstand kleiner oder gleich 12	->	0,5 Punkte	Abstand größer 12	->	0 Punkte
Abstand kleiner oder gleich 1,5	->	4 Punkte														
Abstand kleiner oder gleich 3	->	2 Punkte														
Abstand kleiner oder gleich 6	->	1 Punkte														
Abstand kleiner oder gleich 12	->	0,5 Punkte														
Abstand größer 12	->	0 Punkte														

Number interface ("My current estimate")

Verbleibende Zeit [sec]: 63

Runde 1 von 1. Sie sind **M5** und spielen mit **9 Personen**. Entscheidungs-Regel: **Mehrheit**. Kommunikation: **mit Zahlen**.

Frage: Wie viel Prozent der Erdoberfläche ist Wasser?

Kommunikation mit Zahlen

M1	M2	M3	M4	M5	M6	M7	M8	M9
70.0	75.0	72.0	80.0	71.0	80.0	50.0	67.0	66.0
71.0	73.0	67.0				70.0		

Schätzungen, die Sie noch abgeben müssen: 0

Meine aktuelle Schätzung:

OK

Entscheidung

Sobald jedes Mitglied 10 Zwischen-Schätzungen eingegeben hat können sie Ihre Entscheidung eingeben. Entscheidungsregel: **Mehrheit**

Entscheidungs-Regeln

Individual:
Die Kommunikation dient nur dem Austausch, Sie treffen eine individuelle Entscheidung. Es muss keine Gruppenentscheidung gefunden werden.

Mehrheit:
In Ihrer Gruppe muss sich eine Mehrheit bilden, so dass mehr als die Hälfte der Mitglieder die gleiche Entscheidung eingibt. Gelingt das nicht, gibt es keine Auszahlung.

Konsens:
In Ihrer Gruppe müssen alle Mitglieder die gleiche Entscheidung eingeben. Gelingt das nicht, gibt es keine Auszahlung.

Auszahlungs-Intervalle

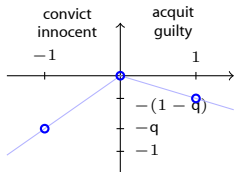
Auszahlungen richten sich nach dem Abstand der Entscheidung zum richtigen Wert:

Abstand kleiner oder gleich 1,5	->	4 Punkte
Abstand kleiner oder gleich 3	->	2 Punkte
Abstand kleiner oder gleich 6	->	1 Punkte
Abstand kleiner oder gleich 12	->	0,5 Punkte
Abstand größer 12	->	0 Punkte

Comparison: Expert Committee and Jury Voting

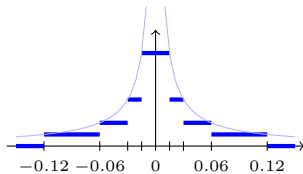
Jury Voting

- ▶ Truth 0 or 1
(innocent or guilty)
- ▶ Decisions 0 or 1
(acquit or convict)
- ▶ Have private signal/opinion



our Expert Committees

- ▶ Truth 0 to 100
- ▶ Decisions 0 to 100
- ▶ Have own knowledge

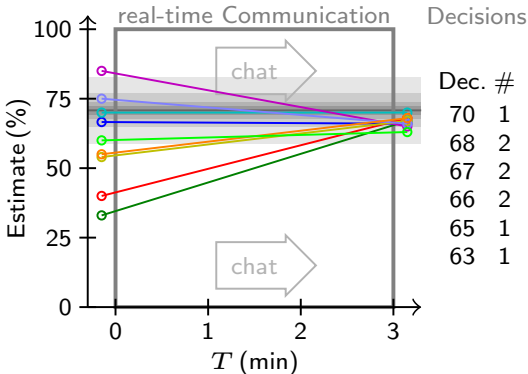


Example: Chat Individual

ID 197 Session 5

Water on earth surface (Q2)

Decision Rule: Individual

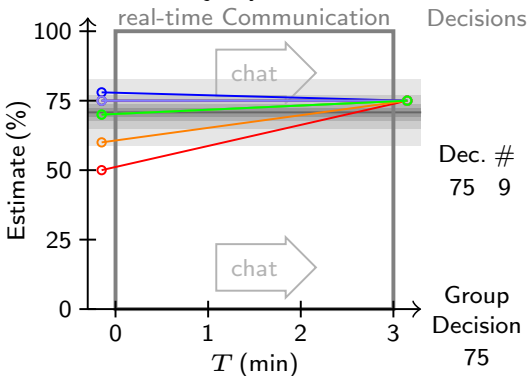


Example: Chat Majority

ID 6 Session 1

Water on earth surface (Q2)

Decision Rule: Majority

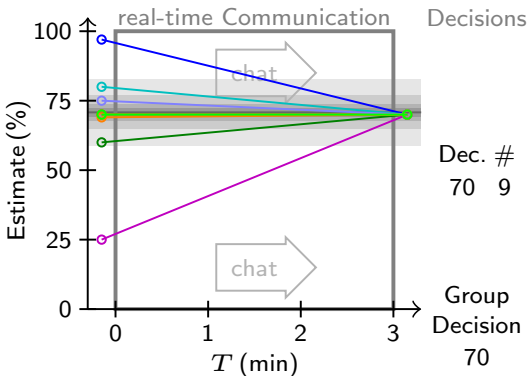


Example: Chat Consensus

ID 150 Session 4

Water on earth surface (Q2)

Decision Rule: Consensus

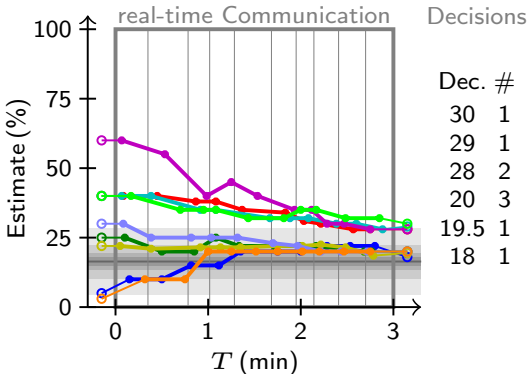


Example: Numbers Individual

ID 165 Session 4

Votes USA at World Bank (Q6)

Decision Rule: Individual

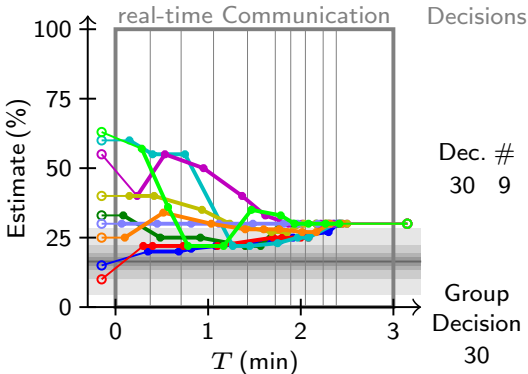


Example: Numbers Majority

ID 213 Session 5

Votes USA at World Bank (Q6)

Decision Rule: Majority

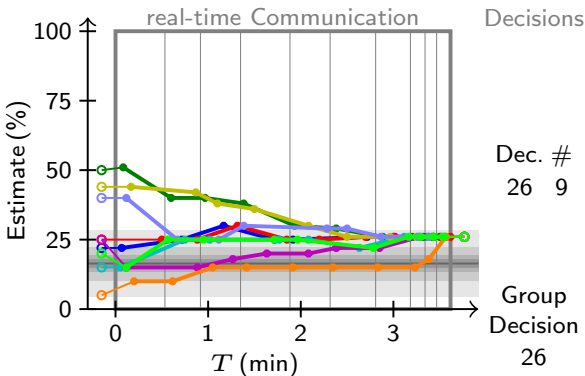


Example: Numbers **Consensus**

ID 22 Session 1

Votes USA at World Bank (Q6)

Decision Rule: Consensus

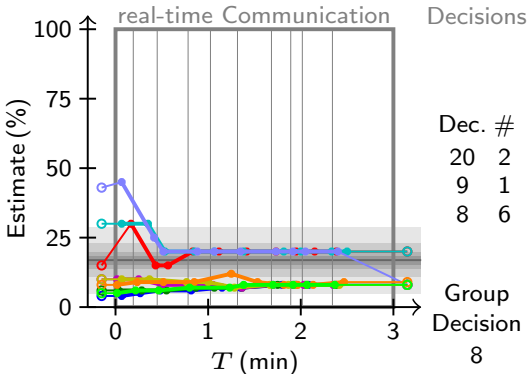


Example: No Unanimity under Majority Rule

ID 41 Session 1

Fat content reindeer's milk (Q11a)

Decision Rule: Majority

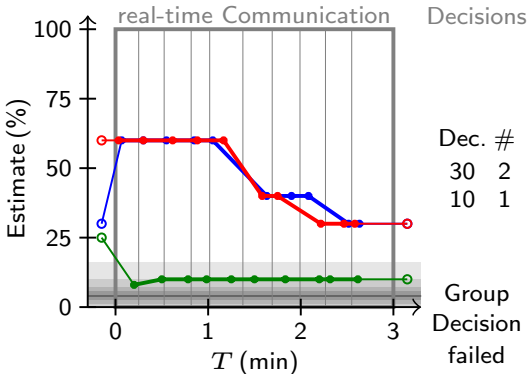


Example: No Decision under Consensus Rule

ID 187 Session 4

Fat content mother's milk (Q11b)

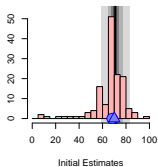
Decision Rule: Consensus



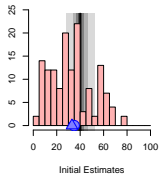
Is there a Wisdom-of-Crowds Effect in Initial Estimates?

Question	Truth	Mean ○	Median △	Subjects better than mean	Subjects better than median	Median Pay. → Pay. median
Q2: Water on earth surface	70.8	68.2	70	34%	1%	€1 ↗ €4
Q4: Workers in agriculture	40.0	34.8	33	23%	23%	€0 ↗ €0.5
Q6: Votes USA at World Bank	16.4	33.2	30	55%	39%	€0 → €0
Q8: Use of own oil Saudi Arabia	19.0	13.5	15	25%	31%	€0.5 → €0.5

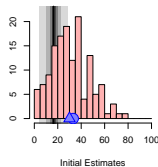
Q2: Water on earth surface



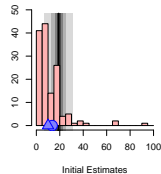
Q4: Workers in agriculture



Q6: Votes USA at World Bank



Q8: Use of own oil Saudi Arabia



- ▶ On average “Yes”, but questions with **population bias** exist.
- ▶ “Median” is slightly more powerful as aggregator of the

Improvement of Payment for Median after Deliberation

Analysis of Questions 2,4,6,8; N = 92 Groups

(Unit: €)	All	Numbers	Chat	3	9
All	-0.17	-0.30	-0.04	-0.35	-0.01
Individual	0.17	-0.04	0.34	-0.39	0.66
Majority	-0.76**	-0.66**	-0.87*	-0.77	-0.75*
Consensus	0.08	-0.19	0.37	0.10	0.06

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

- ▶ Individual and consensus groups: No improvement
- ▶ **Majority** groups: Significant and substantial **worsening**
- ▶ “Chat” better than “Numbers”
- ▶ Only individual groups benefit from larger groups

Improvement of Certainty after Deliberation

Analysis of Questions 2,4,6,8; N = 92 Groups

(Likert-Scale 1–6)	All	Numbers	Chat	3	9
All	0.33***	0.26*	0.40**	0.44**	0.23*
Individual	0.26**	0.15	0.35**	0.29*	0.23*
Majority	0.52***	0.49*	0.55**	0.51	0.52***
Consensus	0.21	0.12	0.31	0.51	0.07

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

- ▶ Average confidence in group decision is **larger** than in individual initial estimates.
- ▶ Effect largest in majority groups, lowest in consensus groups

Conclusion

- ▶ **Majority rule** seems to be a **bad mechanism** for **expert committees** because
 - ▶ Majority decision often **worse** than median of initial estimates
 - ▶ It triggers **most increase of confidence** in quality of decision
- ▶ Consensus decisions might lower the increase of confidence compared to individual decisions
- ▶ Free text communication seems to better improve group accuracy than communication by numbers only

Other Questions

- ▶ There exist questions with large population bias where **deliberation can improve** beyond the WoC-effect.
- ▶ The fact that majority is worse than other rules remains

Adding Methodological Motivations on Experiments

- ▶ Coming from agent-based models and **computer simulation**
“I’ve played a lot with artificial agents, now I also want to play with real agents.”
- ▶ **“Political Engineering”** spirit
“Let us try and test communication modules and decision rules which we might use in reality.”

Appendix: Descriptive Statistics all Questions

	Question	Truth	Mean	Median	IQR	Better Mean	Better Median
1a	Share of people in Asia	60.00	41.31	40.00	25.00	0.40	0.42
1b	Share of people in Americas	13.50	23.65	20.00	12.00	0.62	0.27
1c	Share of people in Africa	15.00	21.30	20.00	17.00	0.51	0.29
2	Water on earth surface	70.80	68.20	70.00	10.00	0.34	0.01
3a	Mass iron in earth crust	4.70	20.30	11.50	20.00	0.56	0.50
3b	Mass aluminium in earth crust	8.10	14.12	10.00	13.25	0.54	0.08
3c	Mass silicium in earth crust	27.70	17.25	10.00	17.00	0.18	0.40
4	Workers in agriculture	40.00	34.83	33.00	26.00	0.23	0.23
5a	Turnout EU parliament UK 1984	32.60	41.51	35.00	32.00	0.35	0.06
5b	Turnout EU parliament FR 1984	56.80	41.58	40.00	29.00	0.33	0.33
5c	Turnout EU parliament DK 1984	52.40	40.38	40.00	38.00	0.29	0.29
6	Votes USA at World Bank	16.40	33.17	30.00	20.00	0.55	0.39
7a	Water content tomato	93.00	77.11	80.00	15.00	0.67	0.38
7b	Water content onion	89.00	53.61	50.00	45.25	0.48	0.48
7c	Water content cucumber	95.00	84.64	88.00	10.00	0.60	0.49
8	Use of own oil Saudi Arabia	19.00	13.47	10.00	15.00	0.25	0.31
9a	Turnout Bundestag DE 2009	70.90	49.74	53.50	29.52	0.52	0.50
9b	Turnout Bundestag DE 1998	82.20	59.94	61.50	17.50	0.60	0.50
9c	Turnout Bundestag 2005	77.70	58.51	60.00	20.00	0.56	0.44
10	Land area with no agriculture	82.00	52.27	50.00	33.00	0.49	0.49
11a	Fat content reindeer's milk	16.90	26.81	23.50	32.00	0.31	0.08
11b	Fat content mother's milk	4.00	29.96	22.50	39.25	0.52	0.50
11c	Fat content water buffalo's milk	8.00	24.99	10.00	34.00	0.62	0.07
12	US presidents from Republicans	40.90	54.76	60.00	16.00	0.32	0.39

Appendix:

Improvement of Payment for Median after Deliberation

Analysis of all Questions; N = 276 Groups

(Unit: €)	All	Numbers	Chat	3	9
All	0.26**	0.15	0.37**	0.20	0.32**
Individual	0.28*	0.32	0.25	0.10	0.45*
Majority	-0.01	-0.25	0.23	0.01	-0.03
Consensus	0.52***	0.39	0.64***	0.48*	0.55**

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1

Appendix:

Improvement of Certainty after Deliberation

Analysis of all Questions; N = 276 Groups

(Likert-Scale 1–6)	All	Numbers	Chat	3	9
All	0.31***	0.28***	0.33***	0.41***	0.21***
Individual	0.33***	0.28**	0.39***	0.42***	0.26***
Majority	0.49***	0.48***	0.5***	0.61***	0.38***
Consensus	0.10	0.07	0.12	0.20	0.00

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1