university of groningen / faculty of behavioural and social sciences / sociology

ICS RUG/UU/RU

Class 2
Testing theories of polarization and cooperation in the Lab

Andreas Flache

Class lecture @
Latin American School and Workshop on
Data Analysis and Mathematical Modeling of Social Science.
SoFiA - SocioFisica Argentina
November, 7-11, 2016
Buenos Aires, Argentina

Part 1: testing models of polarization

- (Some) micromechanisms
- Experiment with online responses ("big data")

What goes really on at the microlevel?

Controlled lab experiments

PLOS ONE

RESEARCH ARTICLE

Discrepancy and Disliking Do Not Induce Negative Opinion Shifts

2016

Károly Takács[1]*, Andreas Flache[2], Michael Mäs[2]

OPEN ACCESS Freely available online

PLOS ONE

Differentiation without Distancing. Explaining Bi-Polarization of Opinions without Negative Influence

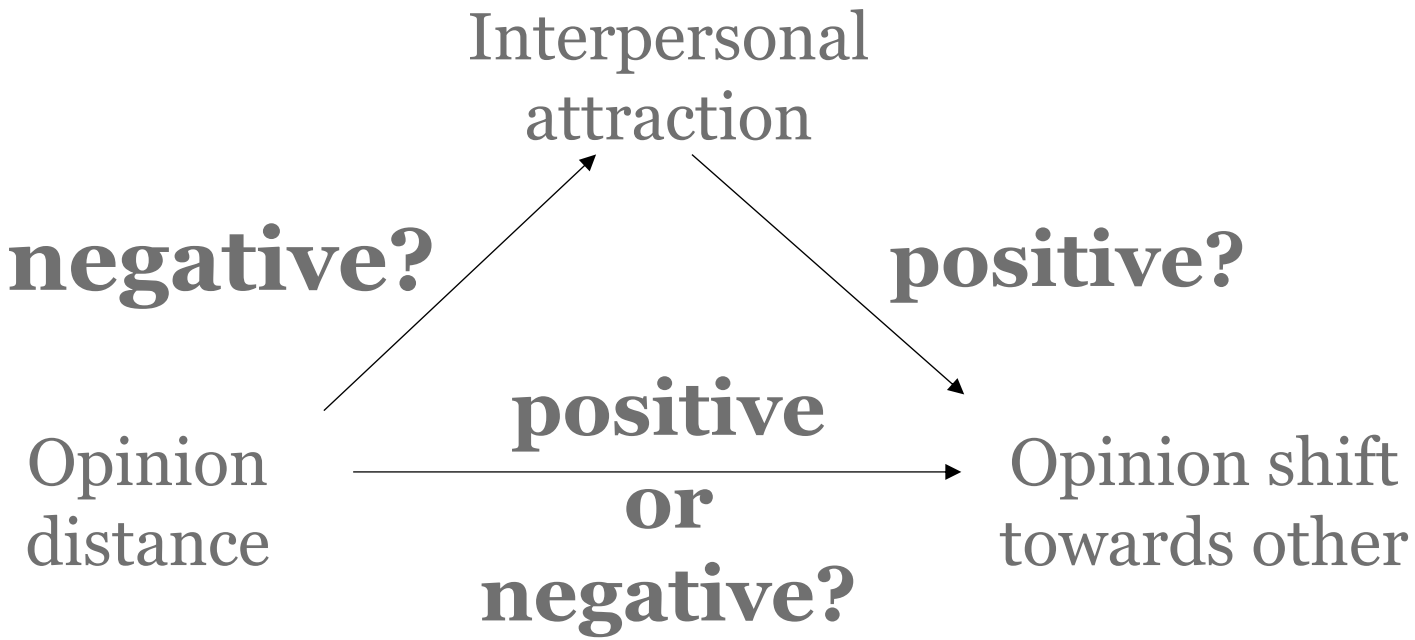
2013

Michael Mäs[1]*, Andreas Flache[2]

1 Chair of Sociology, in particular Modeling and Simulation, ETH Zurich, Zurich, Switzerland, 2 Department of Sociology/ICS, University of Groningen, Groningen, The Netherlands

Focus on "basic building block"

Experiments: main questions



"Stimulus": participant Ego sees opinion of Alter
"Response": change opinion + attraction towards Alter

Why just focus on "basic building block"?

Observed associations in natural data can have many reasons

Examples:

- You move towards a friend's opinion because the friends of your friend influence you.
- You move towards the opinion of someone you interact with because you want to move away from some else's opinion.
- Bidirectional causality of liking and agreement

Basic features of experiments (Takács e.a. 2016)

We conducted a series of 4 experiments with in total 443 subjects.

Overall design:

Measure subjects' opinions on pre-selected issues.

- E.g. "0..100 percent of immigrants who come to the Netherlands for economic reasons should receive a residence permit. "

› Pair subjects with variation distance on opinions and other characteristics.

› Repeated sequence of

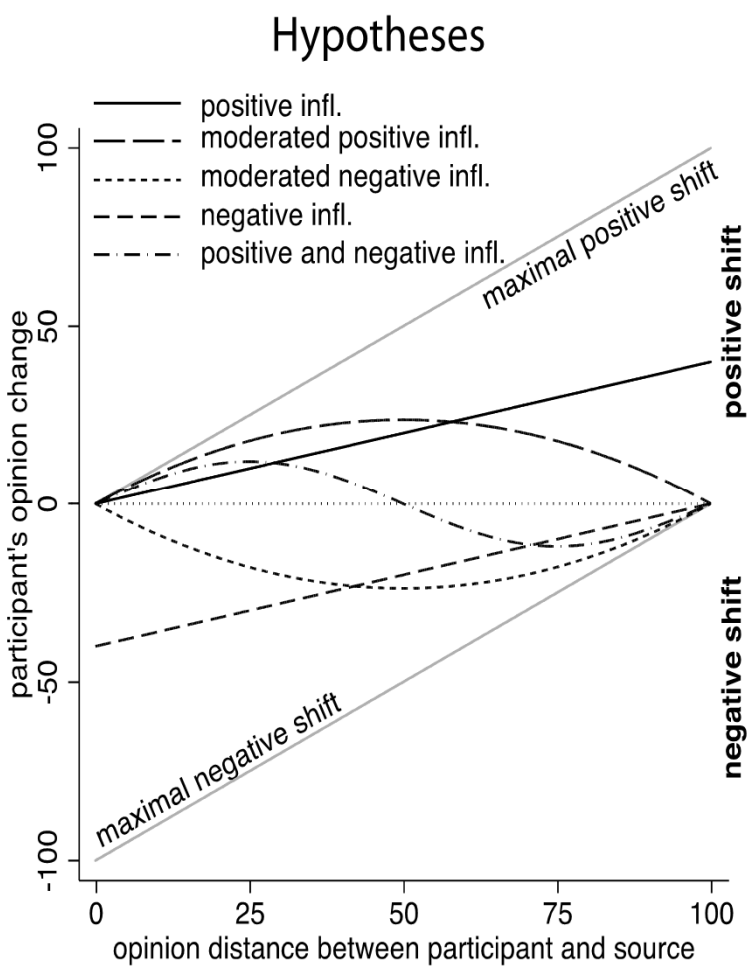
- exposure to others' opinions,
- (exchange messages to influence each other)
- adjust opinions.

› Attractions ("weights") were also measured repeatedly

› In some conditions, we *manipulated initial attraction*

- E.g. dictator games, football support, different moral positions

What to expect? Theory first.



Basic model:

$$\Delta o_{it} = w_{ijt}(o_{j,t} - o_{i,t})$$

Linear positive influence:

$$w_{ijt} = c, \quad c > 0$$

Moderated positive influence:

$$w_{ijt} = 100 - d\left|o_{j,t} - o_{i,t}\right|, \quad w_{ijt} > 0$$

Positive + negative influence:

$$w_{ijt} = 50 - e\left|o_{j,t} - o_{i,t}\right|, \quad w_{ijt} \gtreqless 0$$

Experiment 1

- First, opinions and saliences are measured on 31 issues
- Then “pairs” are formed:
  - An opinion of a person from an earlier session was presented
  - This allowed to conduct these experiments using a web-based questionnaire
- In one session: 7-9 dyadic “pairs”, all with a single issue
- Sequence of: measurement of opinion and attraction updates and controlled messages
- No monetary (!) incentives to change opinions
- No cheating!
- **Induced variation of initial opinion difference**

Opinion issues (preselected in pilot study)

Table S1. Means and standard deviations of original opinions (O) and saliences (S) for issues used in the experiments

Issues	Study 1 (N=89)		Study 2 (N=110)	
	O	S	O	S
1. The warning signs on cigarette boxes should cover 0…100 percent of the box total surface.	43.4 (29.5)	2.04 (.78)	44.5 (29.5)	2.20 (.88)
2. Smoking should be allowed at 0…100 percent of tables in café's.	32.8 (27.5)	1.71 (.57)	23.6 (27.1)	1.75 (.75)
3. The introduction of the euro brings advantages and disadvantages to us. 0…100 percent of all effects are advantages.	51.4 (23.3)	2.16 (.74)	55.6 (24.9)	2.30 (.76)
4. The government should subsidize public transport in 0...100 percent.	69.8 (22.5)	1.56 (.58)	65.7 (22.4)	1.65 (.60)
5. A demonstration needs police protection. Organizers should pay 0…100% of the costs of this.	51.1 (33.1)	2.21 (.67)	44.0 (34.0)	2.35 (.67)
6. 0…100 percent of immigrants who come to the Netherlands for economic reason should receive a residence permit.	34.8 (29.6)	1.83 (.63)	36.3 (29.6)	1.82 (.68)
7. Foreigners who want a residence permit for the Netherlands should pay 0…100 percent of their integration courses and tests.	43.0 (33.5)	1.92 (.79)	42.2 (35.2)	1.96 (.81)
8. 0...100 percent of the streets in the centre of Groningen	32.0	2.15	33.7	2.17

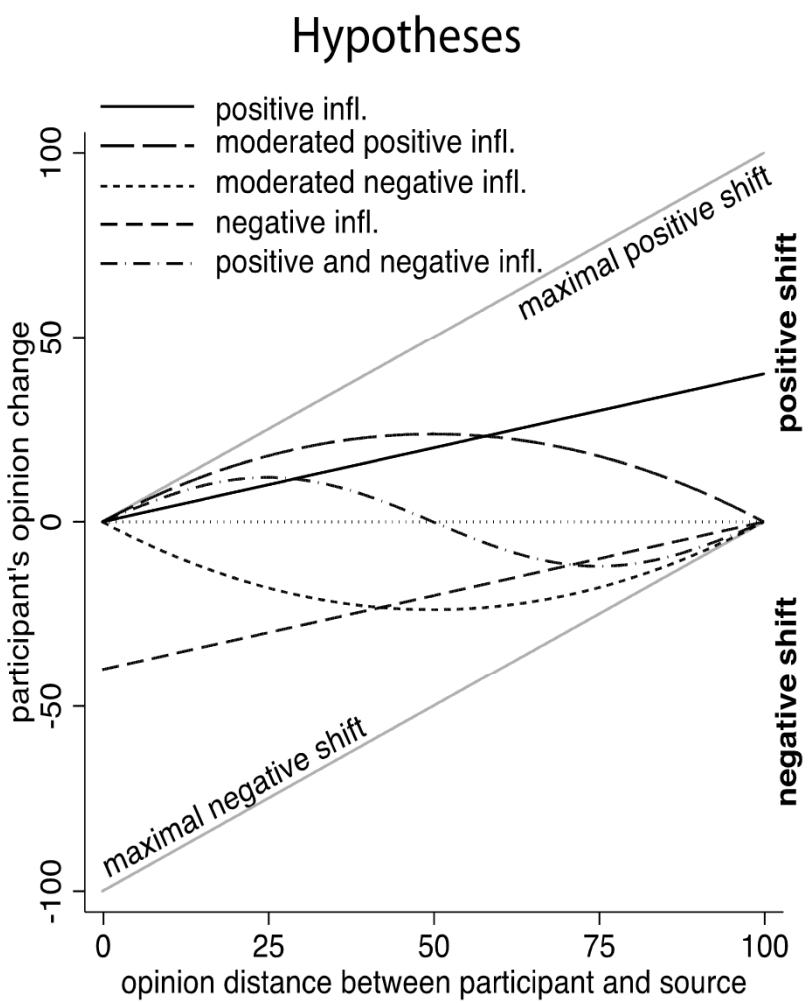
Criteria:

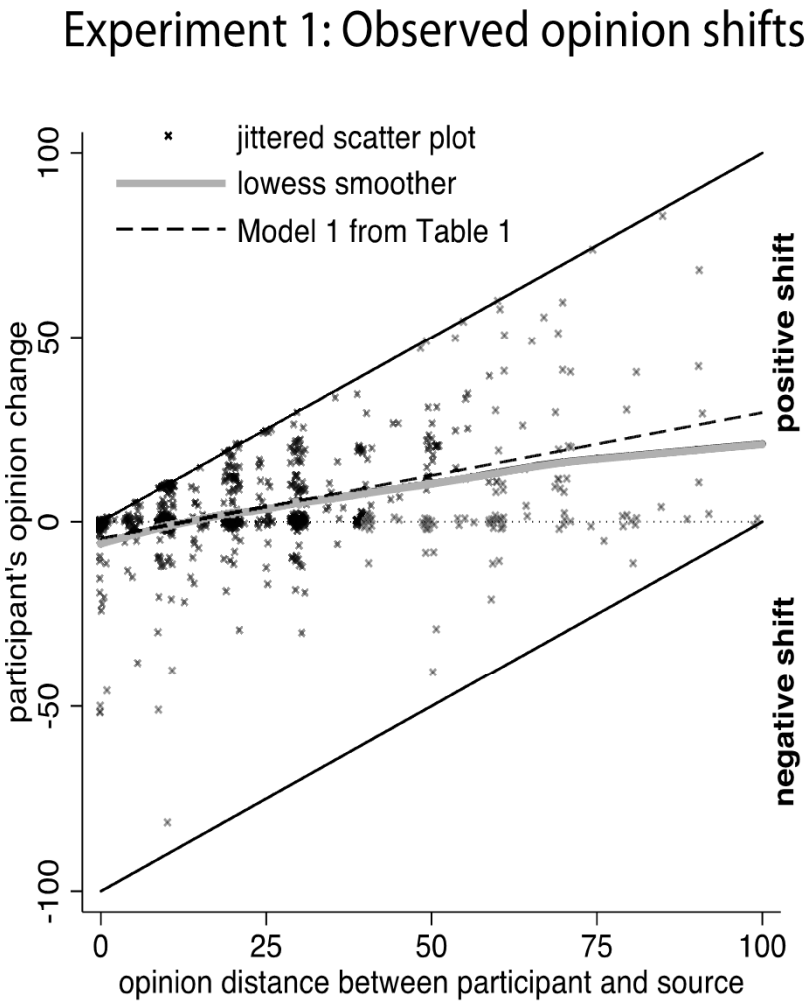
- Enough variation
- Sufficiently salient (but not too salient)

Notes: All issues were measured on a 0…100 percentage scale.

Salience was measured on an ordinal scale: "How important…" with "very important"=1, "important"=2, "unimportant"=3, "very unimportant"=4.

Results for opinion shift





Statistical model of opinion shift

Features

Multilevel random intercept model:

- Nonlinear hypotheses tested with distance^2 and distance^3
- Opinion shifts nested in subjects
- Subject-level individual control variables:
  - Gender
  - Whether subject works
  - Year of study

Statistical model opinion shift

		Fixed effects	
Intercept	-4.35** (1.32)	-4.40** (1.48)	-5.71* (2.53)
Level 1 (observations)			
Distance	0.34 ***(0.09)	0.35* (0.18)	0.34* (0.17)
Distance^2/100	-0.09 (0.14)	-0.12 (0.57)	-0.08 (0.56)
Distance^3/10000		0.02 (0.47)	-0.01 (0.47)
Importance of issue			-0.79 (0.78)
Level 2 (participants)			
Gender (female=1)			1.72 (1.47)
Year of study			0.75 (0.84)
Works (yes=1)			1.48 (1.50)
		Random parameters	
Intercept var. μ_0	21.49***	21.50***	20.87***
level-1 σ^2	159.94	160.21	160.21
Model deviance	4943.56	4942.02	4928.76

Notes. N=617 observations from 89 participants. * p<.05; ** p<.01; *** p<.001.

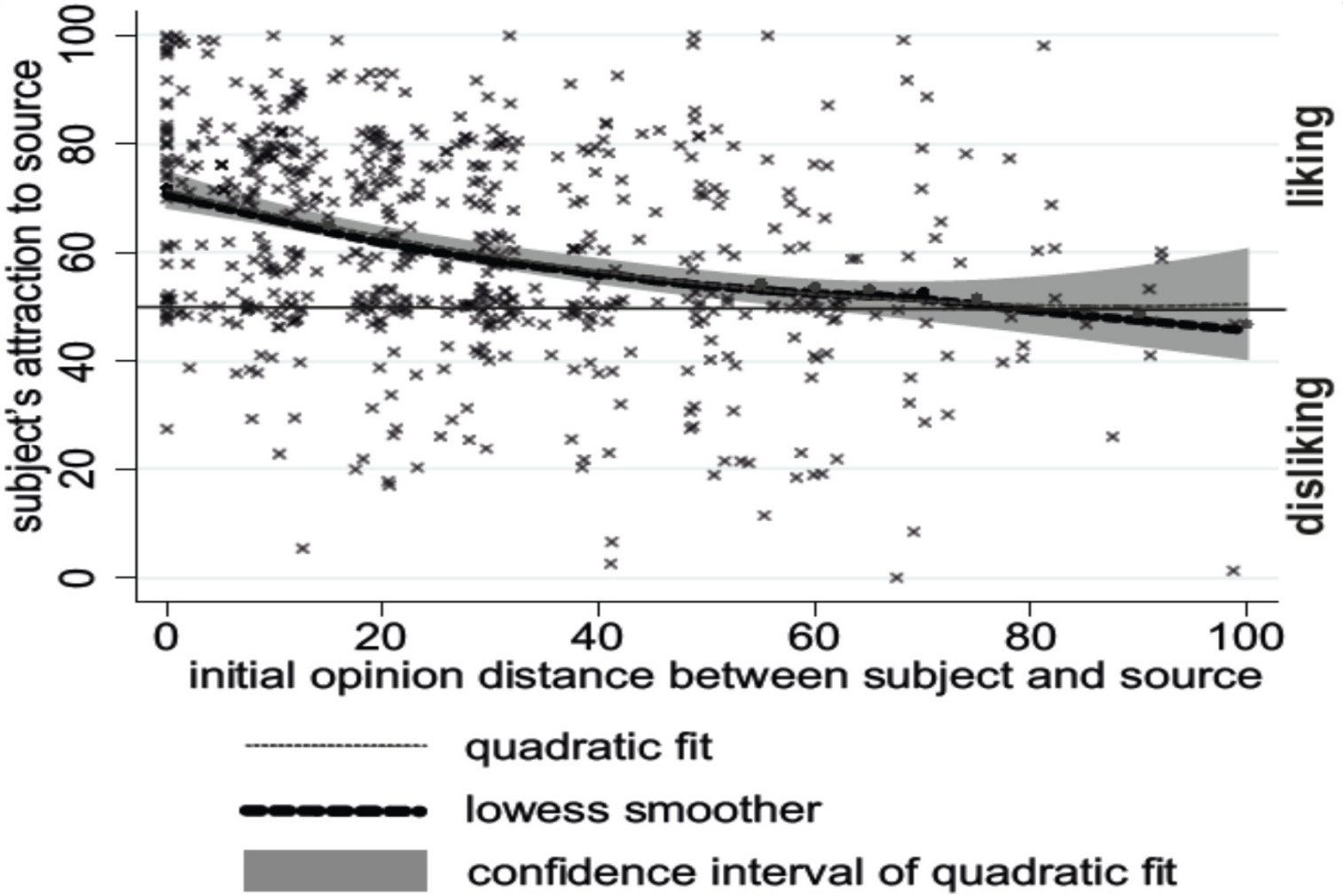
Results statistical model opinion shift

No evidence for non-linear effects of distance

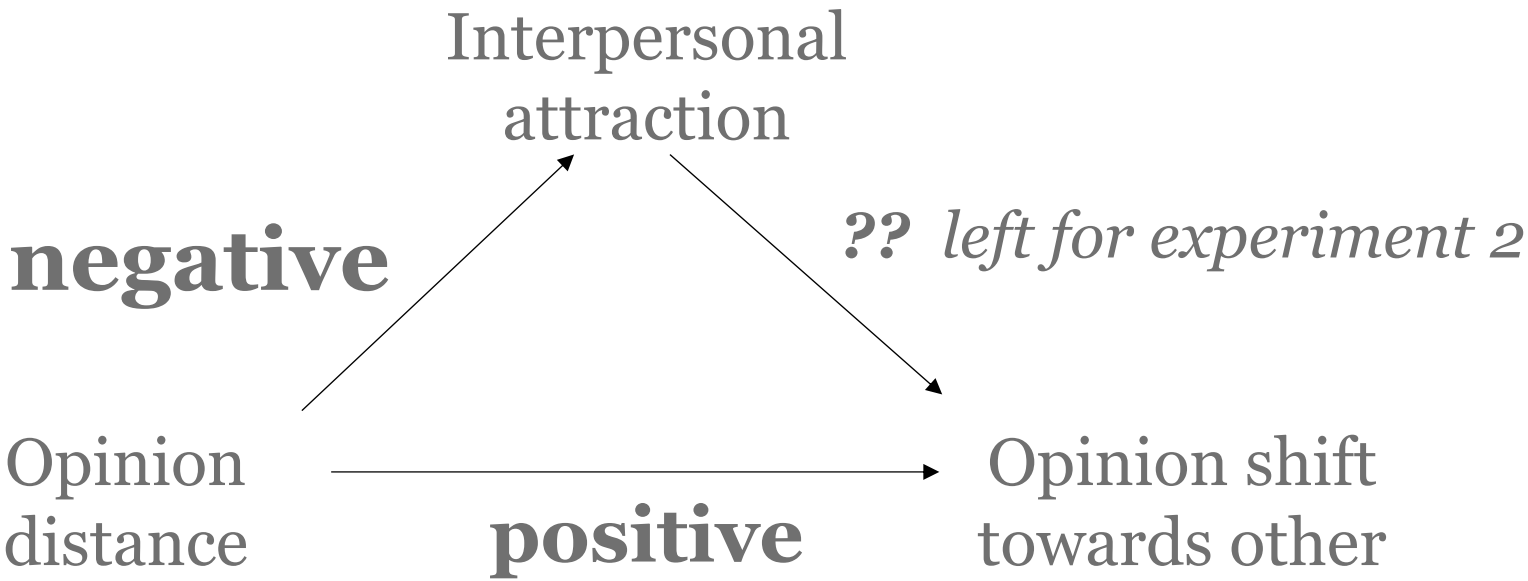
Parameter	Model 1	Model 2	Model 3
		Fixed effects	
Intercept	-4.35** (1.32)	-4.40** (1.48)	-5.71* (2.53)
Level 1 (observations)			
[illegible]	[illegible]	[illegible]	0.34* (0.17)
Distance^2/100	-0.09 (0.14)	-0.12 (0.57)	-0.08 (0.56)
[illegible]0000		0.02 (0.47)	-0.01 (0.47)
Importance of issue			-0.79 (0.78)
Level 2 (participants)			
Gender (female=1)			1.72 (1.47)
Year of study			0.75 (0.84)
Works (yes=1)			1.48 (1.50)
		Random parameters	
Intercept var. μ_0	21.49***	21.50***	20.87***
level-1 σ^2	159.94	160.21	160.21
Model deviance	4943.56	4942.02	4928.76

Notes. N=617 observations from 89 participants. * p<.05; ** p<.01; *** p<.001.

Results for effect of opinion distance on attraction ("homophily" vs "heterophobia")



First experiment: tentative conclusions



Causal effects of attraction on opinion change can not be properly tested because “first stimulus” (opinion alter) affects both opinion participant AND liking of alter at the same time

⇒ **Experiment 2**

E2: Testing underlying mechanisms

- Real dyadic interactions using an experimental software developed for this purpose
- A complex matching algorithm (select 9 issues from 20 for which the best solutions exist):
  - create pairs with low salience-inequality
  - maximize within-individual distance variation
  - create pairs to maximize variation in distances

Subjects

- Mainly students of the University of Groningen, from different faculties
- Subjects are gathered via board advertisements, lecture announcements, and UK advertisements
- Subjects received 8€ for participation + there was a lottery for 200€
- 11 Sessions with 10 participants, total of 110
- Total experimental time: 1 hour
- N(cases) = 920 in 92 issue rounds

Attraction manipulation

Providing information about

- Whether studying at a different faculty
- Whether the other one defected in a PD task
- Whether the other one sent a stigmatizing message

Subjects are only informed about their direct partner

Stigmatizing message:

Participants could chose between sending a stigmatizing or an overwhelmingly positive message to partner.

1. "I am a very nice person. I will do all my best to help you and nobody else in this experiment."
2. "You have to know that I want to do my best in this experiment and I do not care about what you are going to receive."

Manipulation check

Average attraction rating after first stimulus:

- disliking treatment: M=56.57, SD=21.85, 16.3% < 50
- control treatment : M=60.22, SD=20.87; t=2.47, 26% < 50

Multilevel linear regression initial attraction - disliking treatment:

Estimated effects (robust standard errors in brackets)

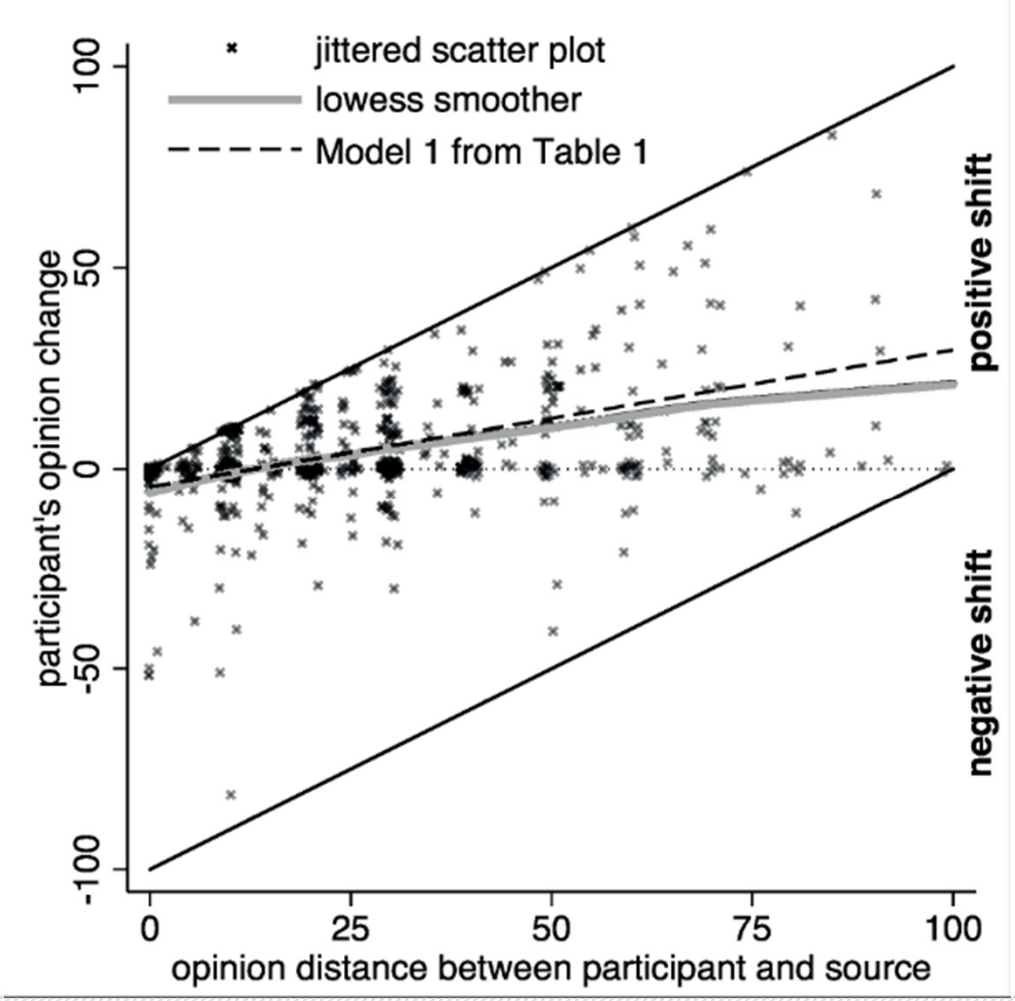
60.57 (1.80)*** (intercept)

– 7.01 (1.89)*** (*PairDefected*)

–6.02 (1.73)*** (*Stigmatized*)

– 1.72 (1.39) (*SameFaculty*).

“Discrepancy and Disliking Do Not Induce Negative Opinion Shifts”



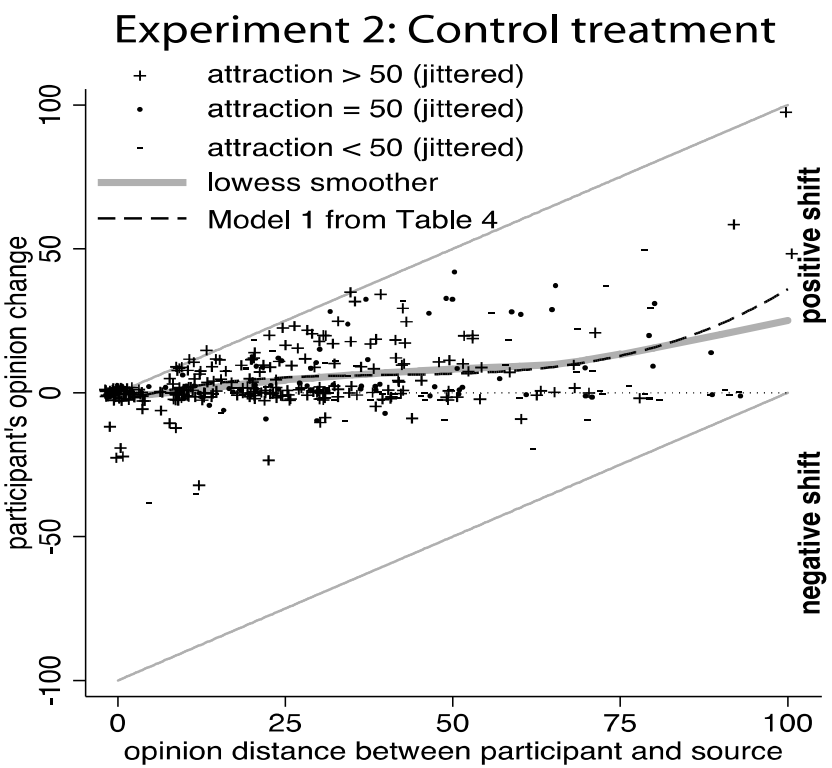
Experiments show:

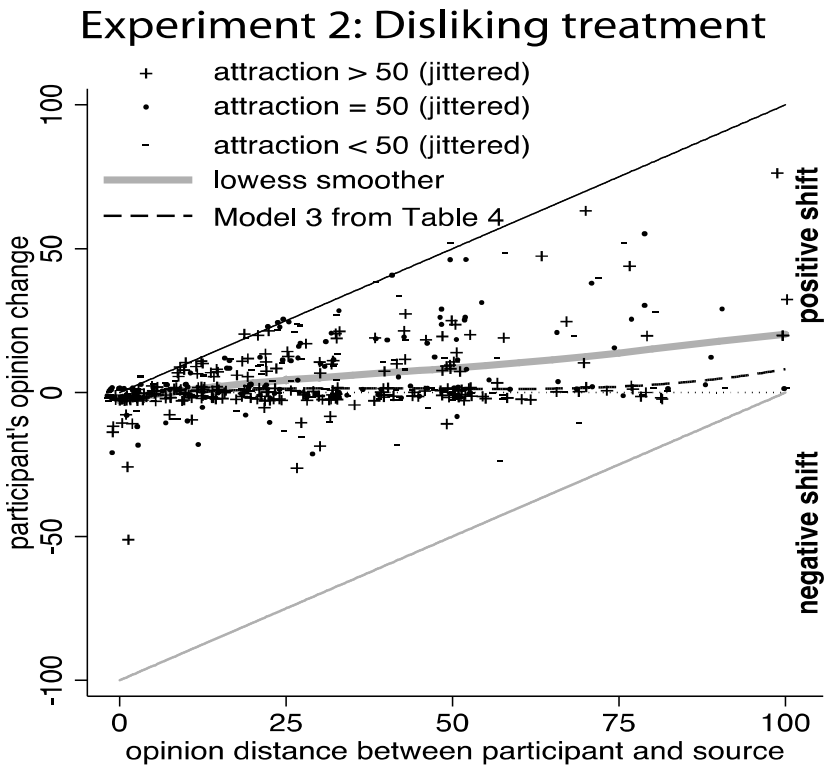
Influence mainly positive

No more negative influence if large disagreement

Takács, Flache & Mäs
Plos One 11(6): e0157948.
doi:10.1371/journal.pone.0157948

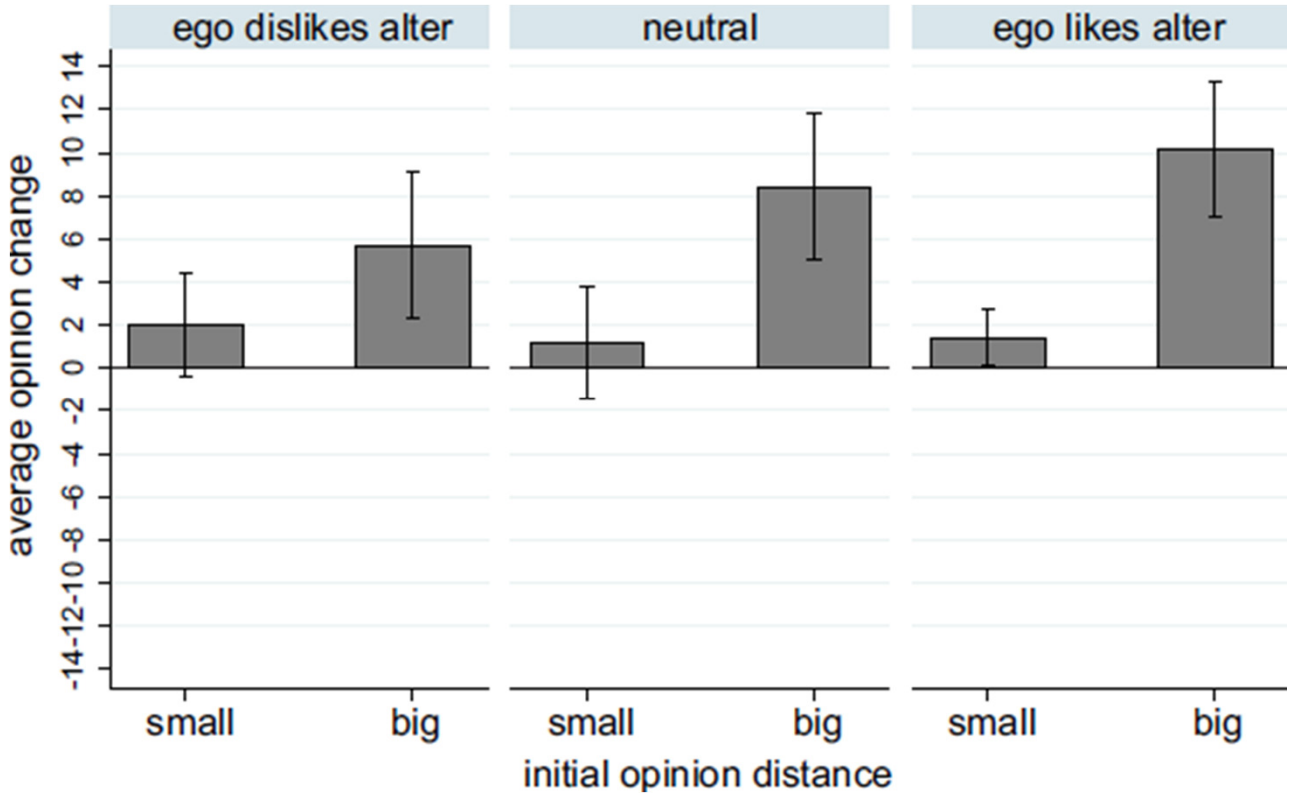
Did induced disliking change influence?





- Induced disliking slightly reduces positive influence
- No evidence that induced disliking elicits negative influence (tested in multi-level regression)

Interaction opinion distance and disliking



Effect of (induced) initial disliking on direction of influence

⇨ No evidence for negative influence in lab.

Tentative conclusions

Best model to explain effects of opinion distance on opinion shift:

› Linear positive influence, proportional to distance

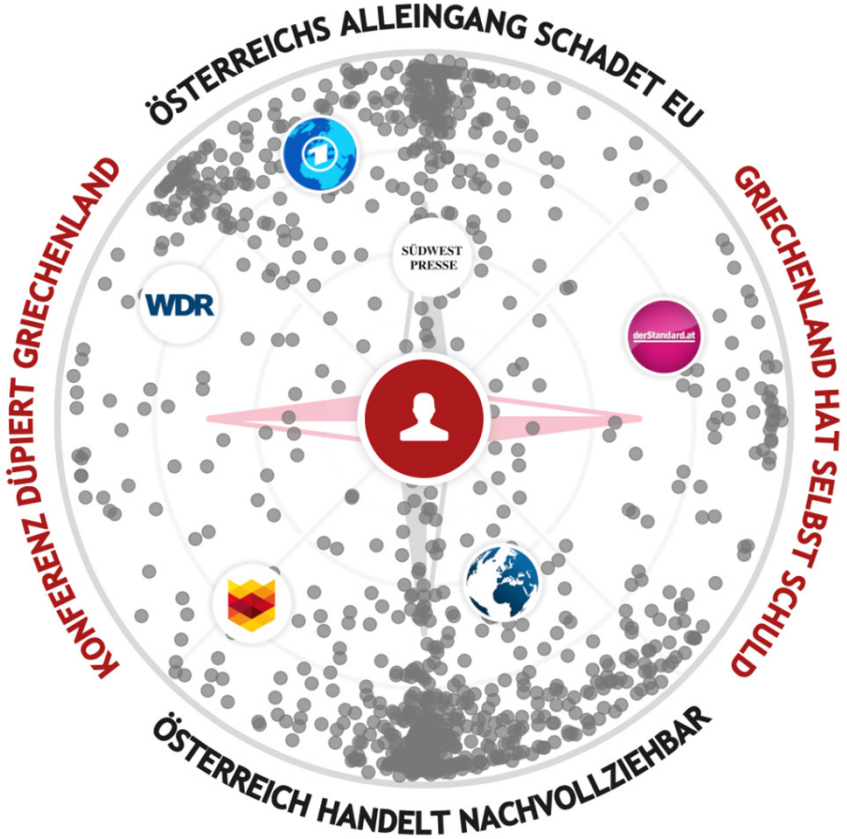
⇒Clear tendency towards “compromise” when opinions differ

There is evidence for (some) “striving for uniqueness”

There is evidence that people dislike others more who differ more from them.

But even when we induced disliking, this did not elicit negative influence.

Social influence and polarization on online news sites

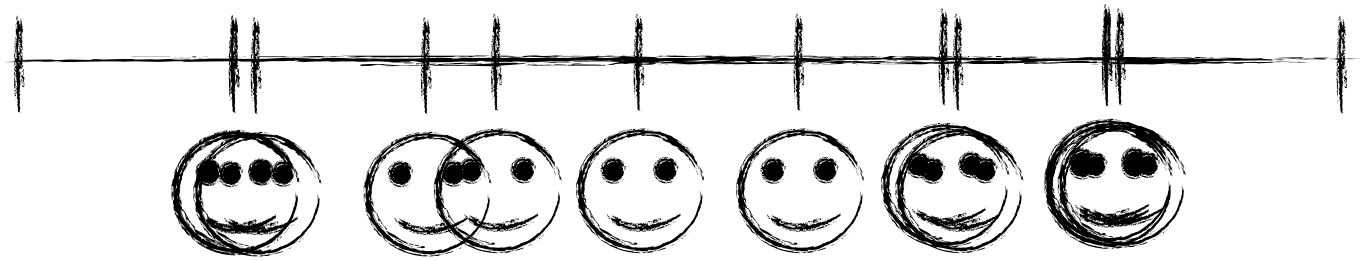


Michael Mäs
Bernhard Clemm von Hohenberg
Bary Pradelski

university of groningen

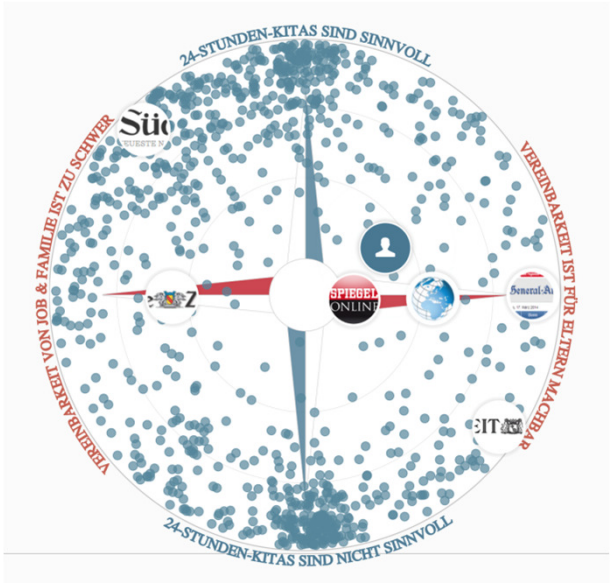
ICS
RUG/UU/RU

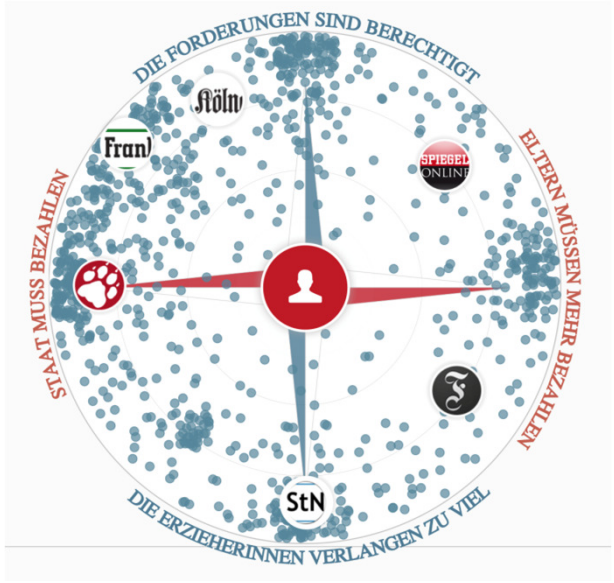
What form of influence is present?

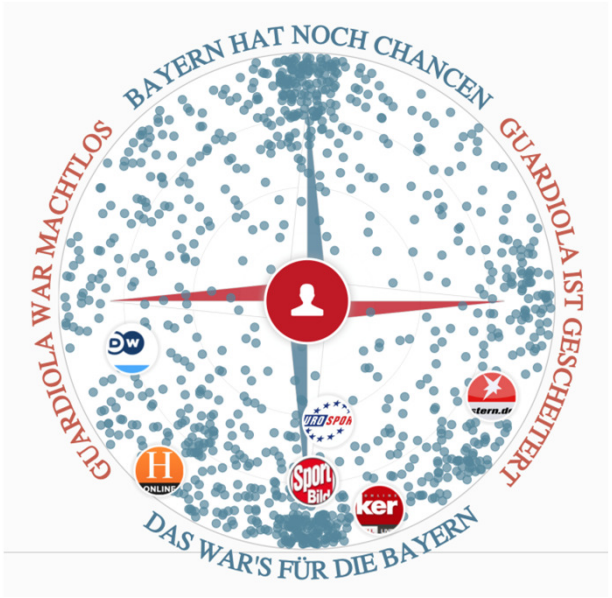


Candidates:
- positive social influence
- moderated positive social influence
- negative social influence
- reinforcement
- striving for uniqueness

THREE EXAMPLES







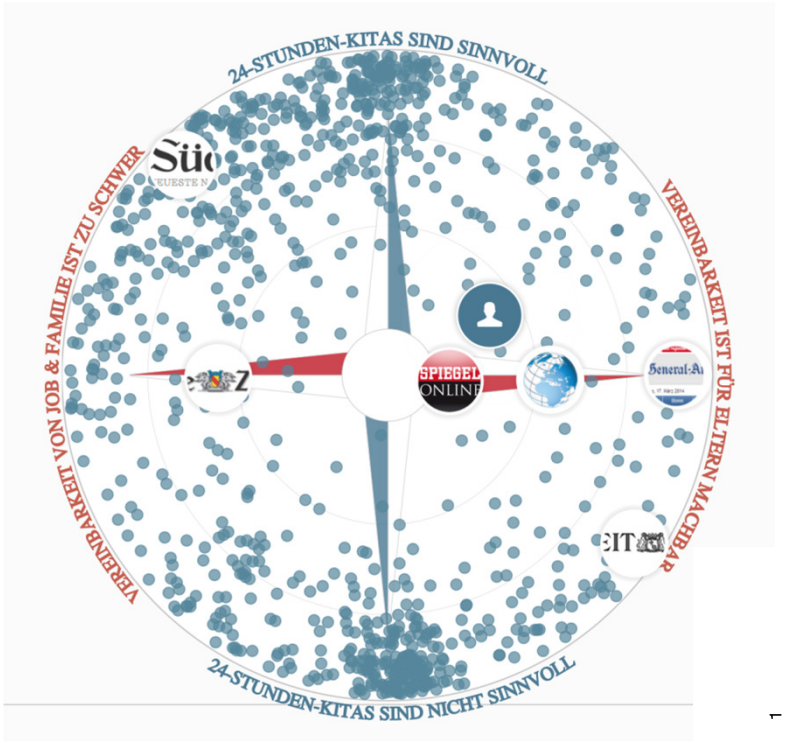
Knowing the distribution of votes, what can one conclude about social influence?

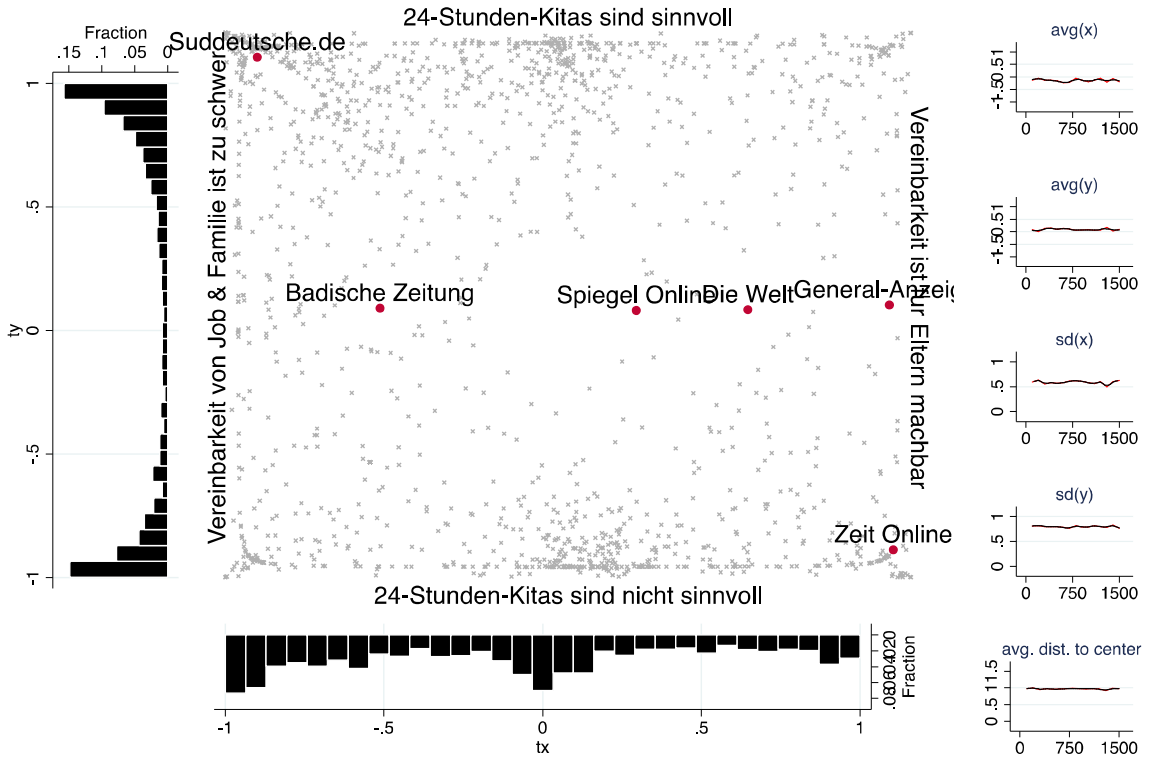
Knowing the dynamics of votes, what can one conclude about social influence?

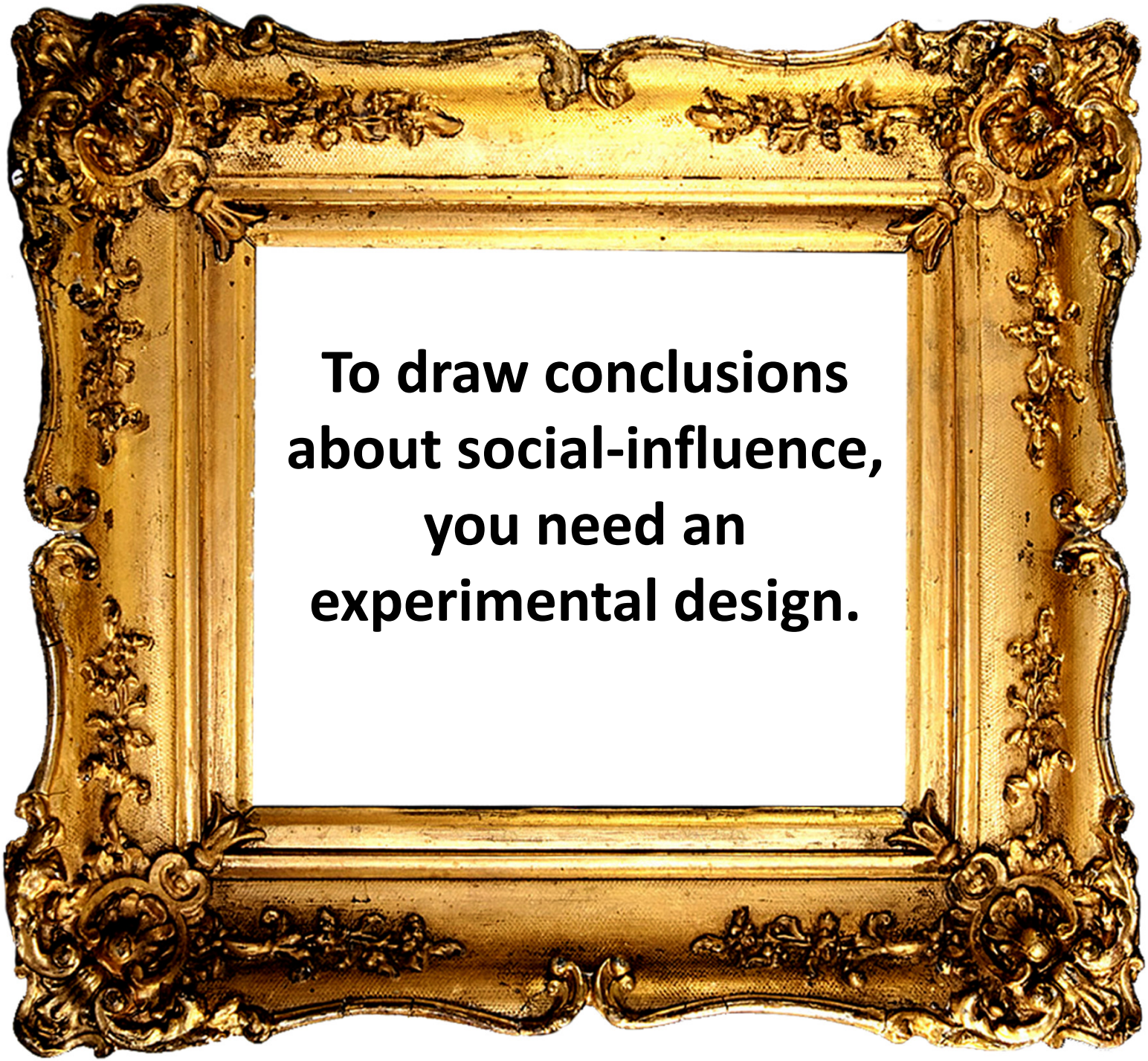
Example: If variance increases, does this show that there is negative influence?
No, it could also be positive influence by extremists.

Example: If variance decreases, does this show that there is positive influence?
No, this might result from self-selection.

Example: If there is no dynamic, does this show that there is no influence?
No, the system might have reached equilibrium


24-STUNDEN-KITAS SIND SINNVOLL
VEREINBARKEIT IST FÜR ELTERN MACHBAR
24-STUNDEN-KITAS SIND NICHT SINNVOLL
VEREINBARKEIT VON JOB & FAMILIE IST ZU SCHWER
SPIEGEL ONLINE


24-Stunden-Kitas sind sinnvoll
24-Stunden-Kitas sind nicht sinnvoll
Vereinbarkeit von Job & Familie ist zu schwer
Vereinbarkeit ist für Eltern machbar
Suddeutsche.de
Badische Zeitung
Spiegel Online
Die Welt
General-Anzeig
Zeit Online
Fraction
ty
tx
avg(x)
avg(y)
sd(x)
sd(y)
avg. dist. to center


To draw conclusions
about social-influence,
you need an
experimental design.

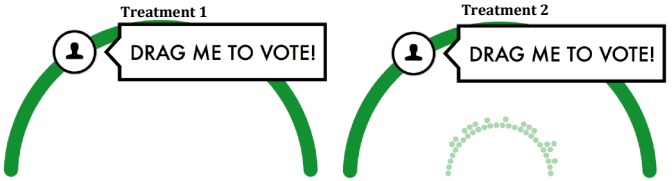
The field experiment

Should there be a female-pilot quota?

Click or drag to take a stance and share your opinion!

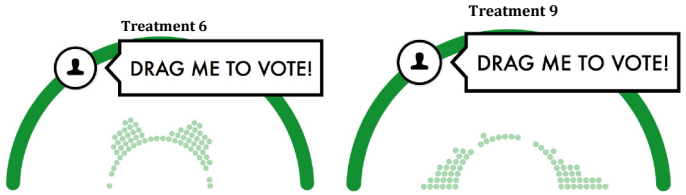
Experimental treatments

Baseline treatments



Single-peaked distributions

Double-peaked distributions



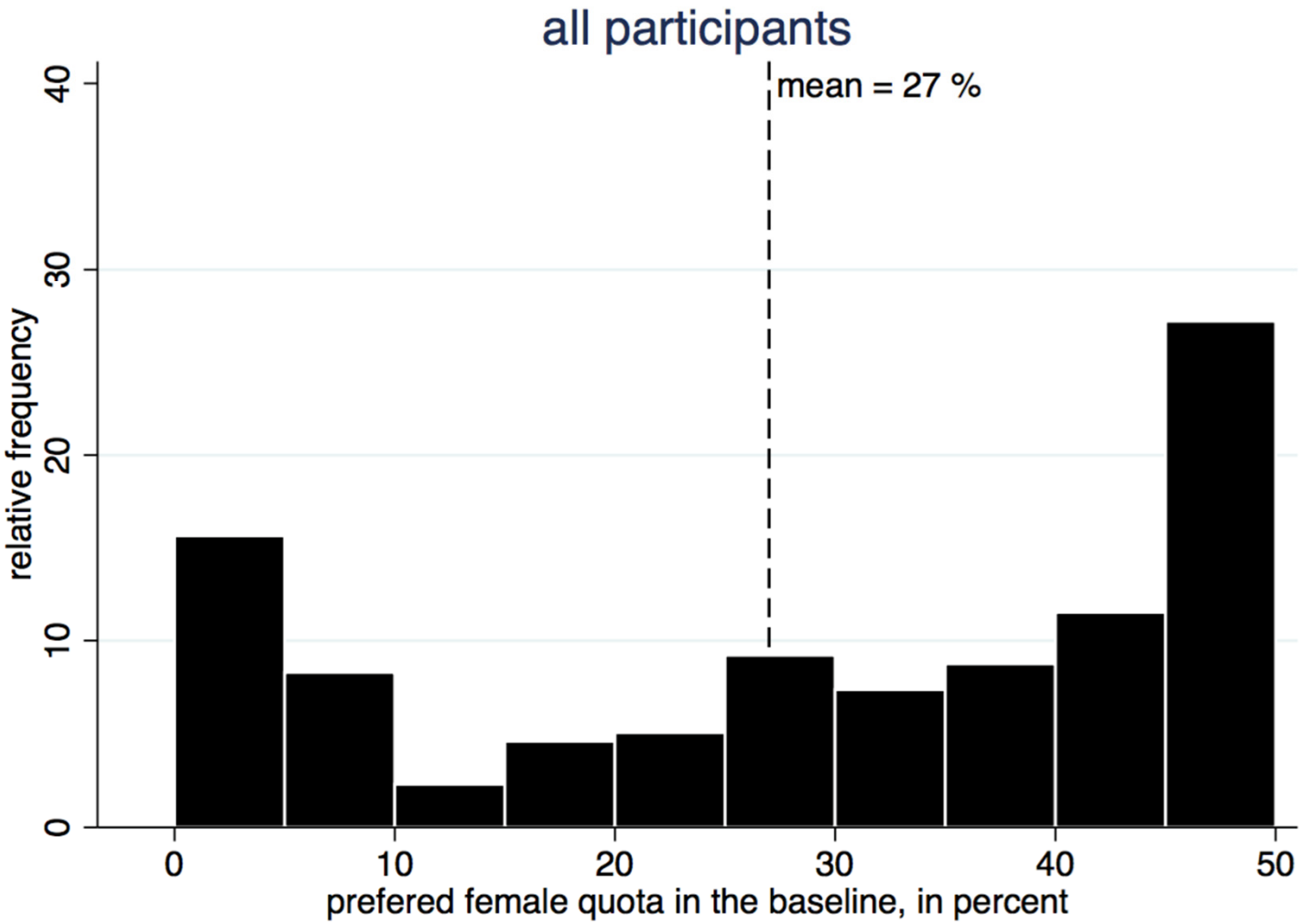
Advantages:

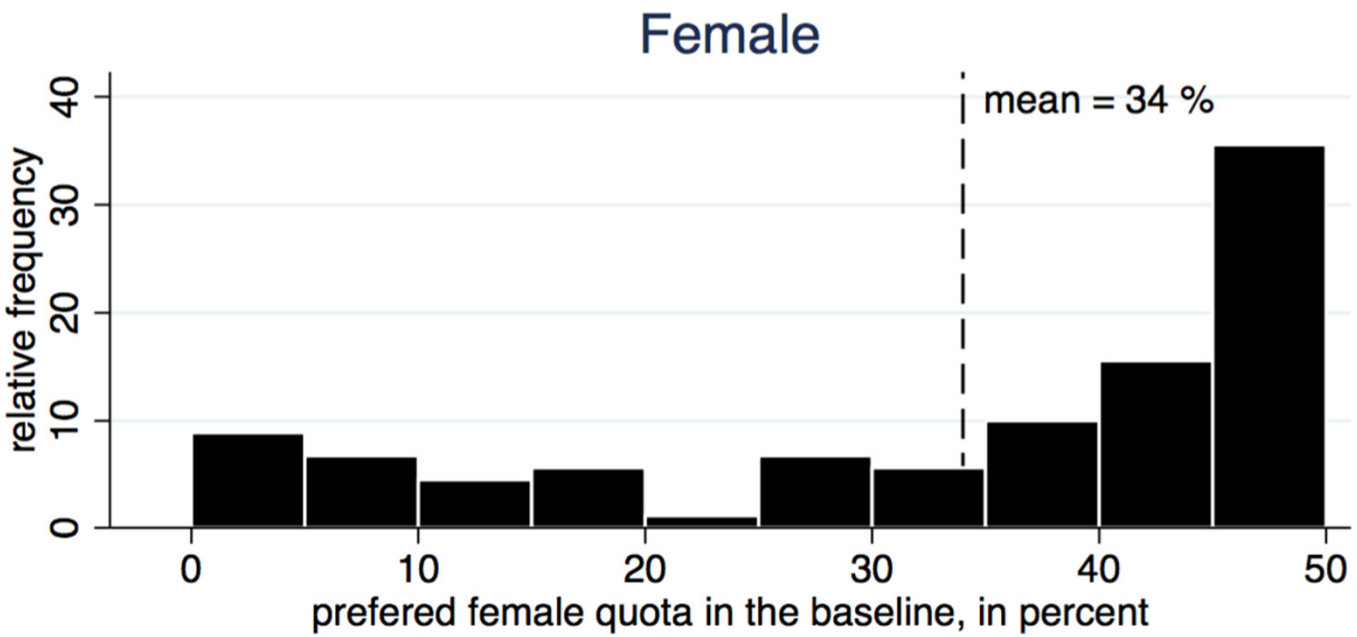
1. Informed consent / no deception
2. We could add a survey
   - 1 week later
   - response rate: 0.8
3. We could compensate
   - no self-selection
   - $ 0.75 + $ 0.20

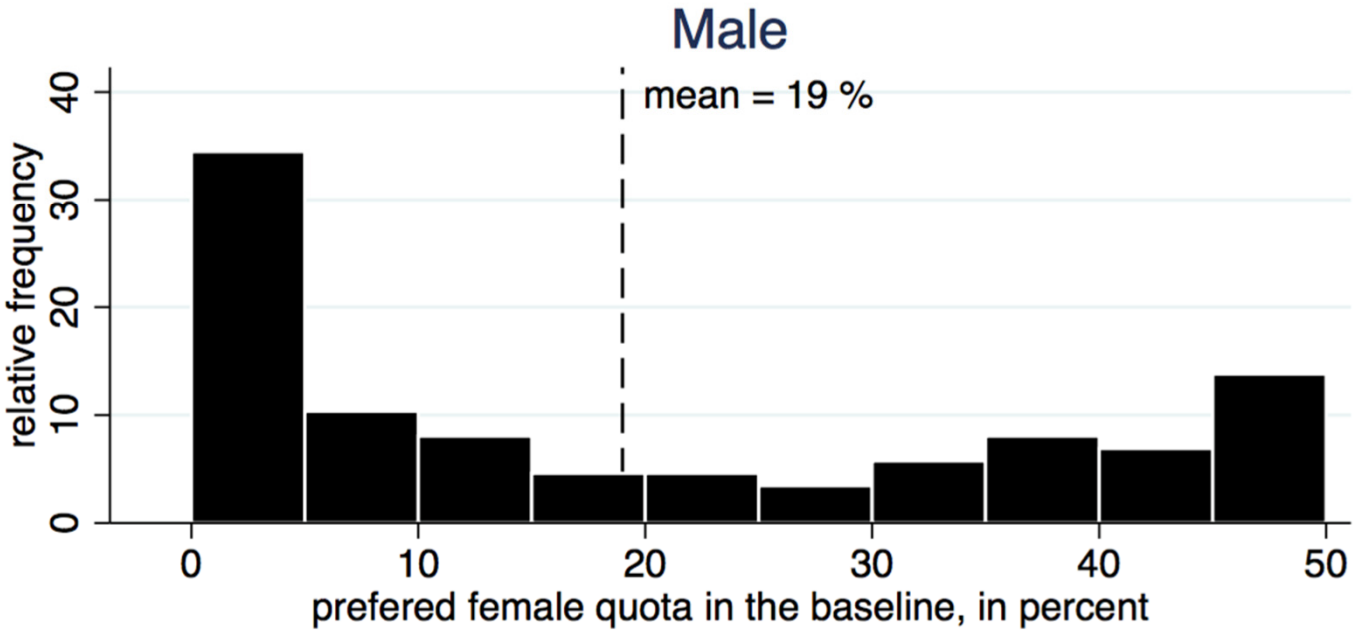
Variable	Relative (%)	Absolute
Female	50.2	1,630
Male	49.8	1,618
Liberal	55.1	1,790
Moderate	19.5	632
Conservative	25.4	826
High school graduate and below	35.2	1,144
College graduate, no higher	49.3	1,600
Higher than college graduate	15.5	504

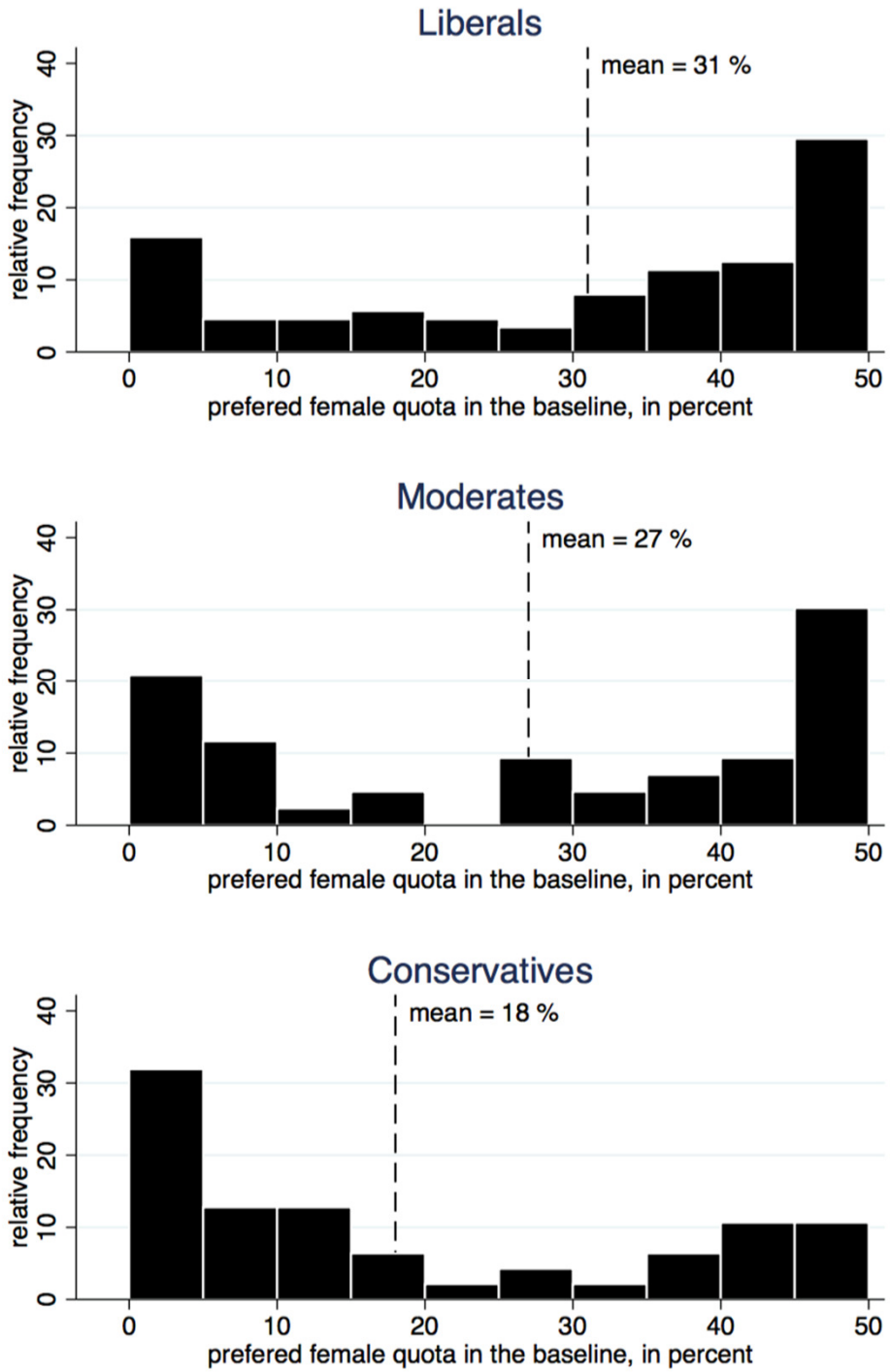
Variable	Mean	St.D.
Age	36.1	11.9

Ratings without influence




Female
mean = 34 %
relative frequency
0
10
20
30
40
0
10
20
30
40
50
prefered female quota in the baseline, in percent


Male
mean = 19 %
relative frequency
0
10
20
30
40
0
10
20
30
40
50
prefered female quota in the baseline, in percent


Liberals
mean = 31 %
relative frequency
0
10
20
30
40
0
10
20
30
40
50
prefered female quota in the baseline, in percent
Moderates
mean = 27 %
relative frequency
0
10
20
30
40
0
10
20
30
40
50
prefered female quota in the baseline, in percent
Conservatives
mean = 18 %
relative frequency
0
10
20
30
40
0
10
20
30
40
50
prefered female quota in the baseline, in percent

How to measure opinion shifts?

Subjects rated their opinion only once, after being exposed to the experimental treatment

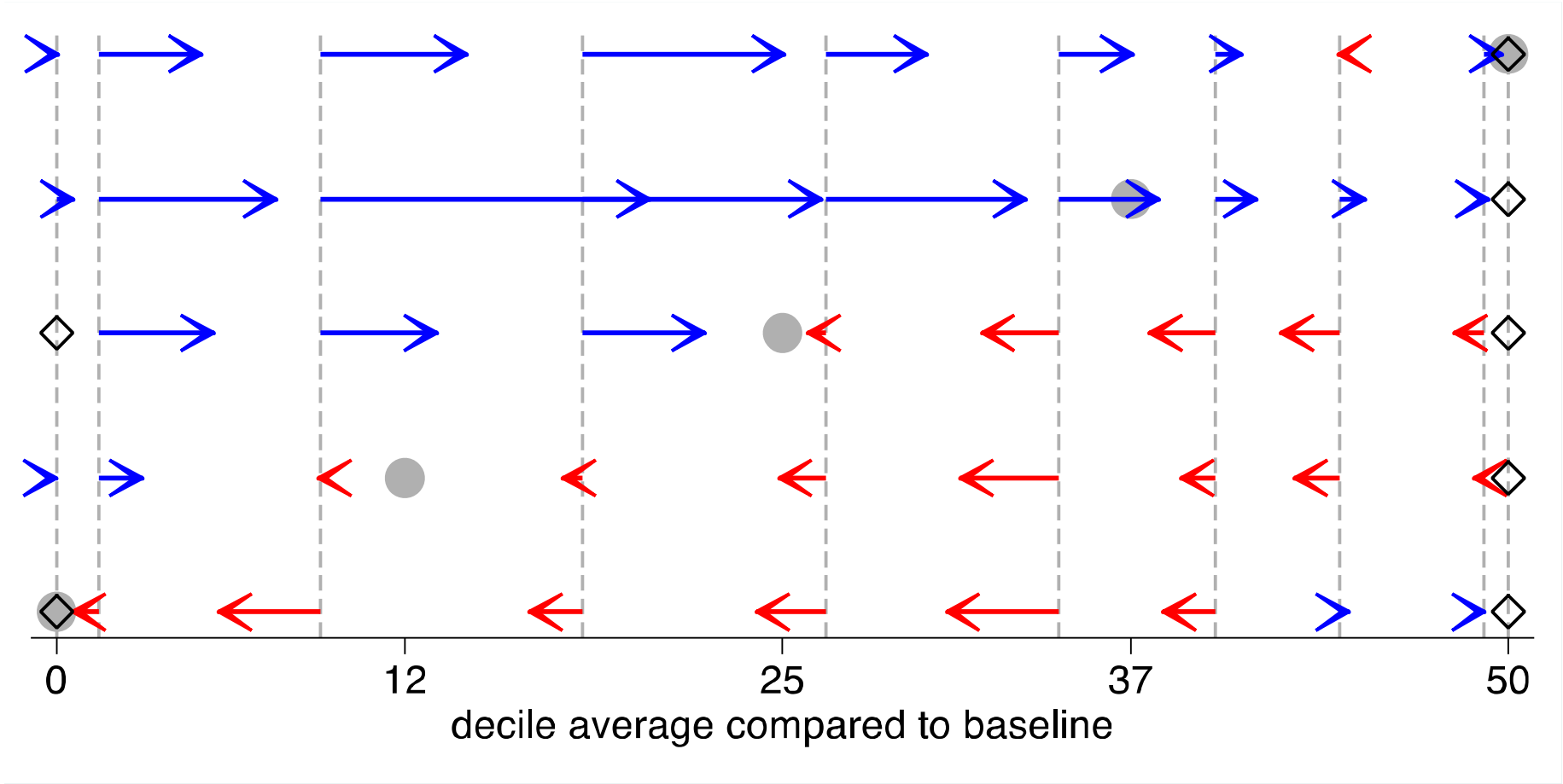
3 survey characteristics explained 40% or rating variance in baseline

- Gender
- Political orientation
- General view on quotas

Prior opinion = median of baseline subjects with the same survey characteristics

No treatment effects on approximated prior opinions and the three matching variables

Treatment effects



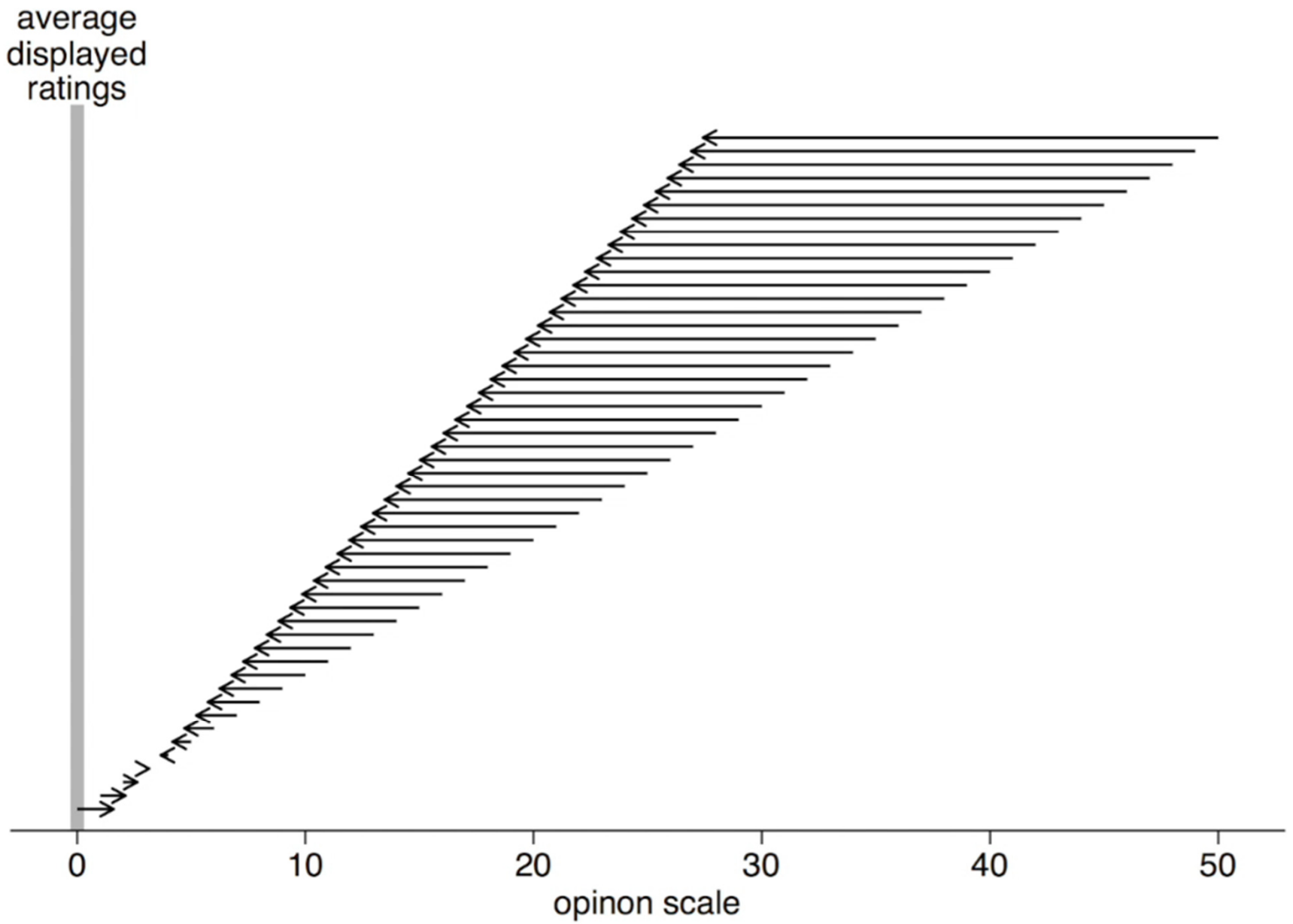
Shifting the peak of the displayed distribution by one percentage point to the right resulted on average in a shift in the rating by 0.1 percent (t = 3.18).

Test of the micro model

Table 3: Linear regressions for social influence model

	Regression 1	Regression 2	Regression 3	Regression 4	Regression 5	Regression 6
dist	0.465				0.460	0.483
	(0.029)**				(0.054)**	(0.029)**
$(dist)^2$		-0.007			-0.001	
		(0.003)**			(0.006)	
$(dist)^3$, standardized			4.042		0.388	
			(0.316)**		(0.656)	
$(dist)^4$, standardized				-1.849	0.485	
				(0.486)**	(0.957)	
constant					1.586	1.596
					(0.736)*	(0.500)**
R^2	0.19	0.01	0.13	0.01	0.20	0.20
N	1,106	1,106	1,106	1,106	1,106	1,106

* $p < 0.05$; ** $p < 0.01$


average
displayed
ratings
0
10
20
30
40
50
opinon scale

Tentative conclusions online experiment

Best model to explain effects of estimated opinion distance to peak of distribution on estimated opinion shift:

- Linear positive influence, proportional to distance

⇒Tendency to compromise between own opinion and "group mean"

After a while, displayed distribution can be in equilibrium.

No change does thus not proof "no influence".

Again: no evidence of "negative influence".

Is this the end of the "negative influence hypothesis"? Surely not …

Under what conditions could there be differentiation from disliked others?

- Social categorization into "us" vs "them" has been shown to evoke negative perceptions of outgroup (e.g. Hogg ea, 1990)

⇒Our "attraction manipulations" focused on the interpersonal level. Future experiments could highlight intergroup differences

⇒Online experiments did not show anything about characteristics of other "voters". This is different e.g. on social media sites.

⇒Test in social media experiments how people respond to messages from sources with different characteristics.

Part 2: testing models of cooperation

- (Some) micromechanisms
- (Some) structural predictions

university of groningen / faculty of behavioural and social sciences / sociology

ICS
RUG/UU/RU

The weak side of informal social control

Counter-reward and counter-punishment in collective good games

Andreas Flache

Dieko Bakker

Jacob Dijkstra

Michael Mäs

Collective action and informal social control

- The free rider problem in collective good production
  - Contribution is costly, but free riders are hard to exclude

  ⇨ groups may fall far short of optimal provision

- One of the solutions is informal social control
  - Reward contributors and / or punish free riders
  - Shunning, approval, ostracism…

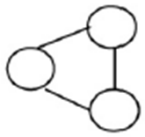
- The "second order free rider problem"
  - Provision of social control imposes in itself a collective action problem.

Endogenous solutions…

- Altruistic punishment (Fehr & Gächter 2002)

  "free riding causes strong negative emotions … most people expect these emotions"… "emotions trigger punishment"

- Network closure and low cost selective reward (Coleman, 1990)



"An expression of encouragement or gratitude for anothers' action may cost the actor very little but provide a great reward for the other"

- Emotional dependence on cohesive group (Homans 1974)

  Group members reward one another with expressions of approval. "Ostracism is the penalty for failing to conform". The more cohesive the group, the stronger the pressure.

…and a new problem?

- Most groups exist over longer time periods

  What happens when informal social control is only one part of a network of ongoing exchange relations between group members?

⇨ not only free riders depend on punishing group members for obtaining social approval, but also vice versa.

⇨ "Counter-punishment" (e.g. Nikoforakis 2008) or 'counter-reward' (e.g. Flache & Macy, 1996) also become possible strategies

How social control fails in cohesive groups: student group assignments

Quotes from Groningen students about group assignments:

"It seemed so nice and cosy with four friends in a work group. But in the end, you are more inclined to take it easy. You expect that your friends will shoulder the burden for you"

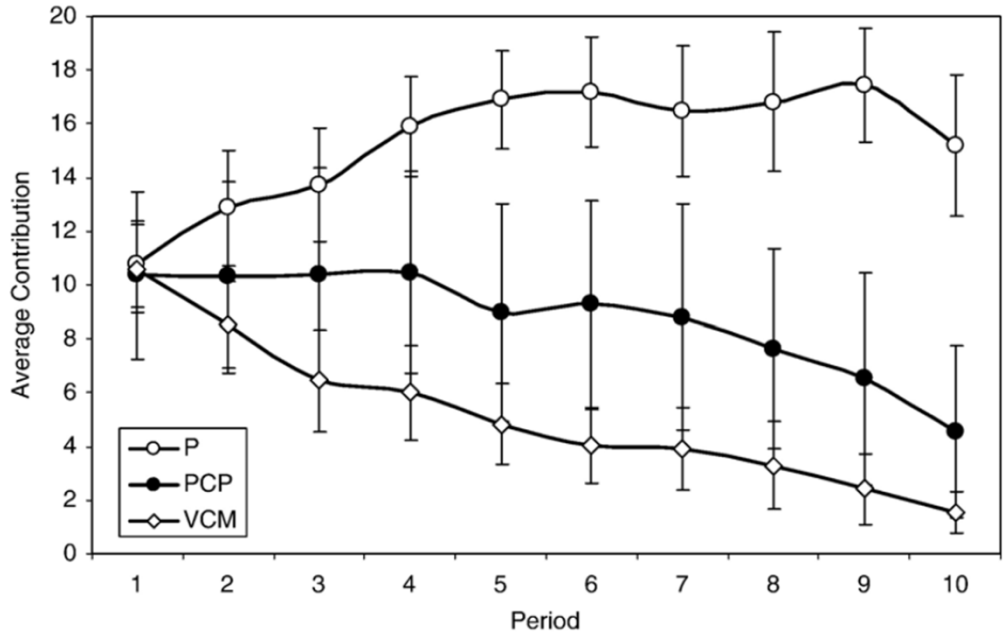
"It is much easier in a group of friends to come with some poor excuse if you did not show up once more"

"In assignment groups, students criticize each other practically never for lack of activity, because students who are also friends just will not let each other down." (quote from a teacher)

University newspaper September 2001 (translation by AF)

Punishment and counter-punishment in public good game (Nikiforakis 2008)

Contributions → Punishment → Counter-punishment



Peer sanctioning: Reward vs. Punishment

› Second order free rider problem? Solutions

- Reward: exchange

  (e.g. Homans, Coleman, Flache & Macy, Willer …)

- Punishment: emotions, “altruistic punishment”

  (Fehr & Gächter, Gintis, …)

Counter-sanctioning in ongoing exchange: a problem for peer sanctioning institutions?

Ongoing exchange process.

⇨ Free riders depend on other group members, but also vice versa.

⇨ *"counter-reward"*

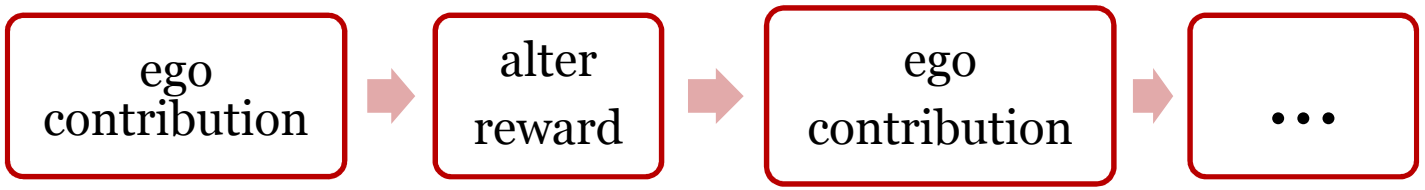
(e.g. Flache & Macy, 1996)

⇨ *"counter-punishment"*

(e.g. Nikoforakis 2008)

Counter-reward

Effective peer sanctioning (without counter-reward)



Ineffective peer sanctioning (with counter-reward)

Counter-reward

Effective peer sanctioning (without counter-reward)

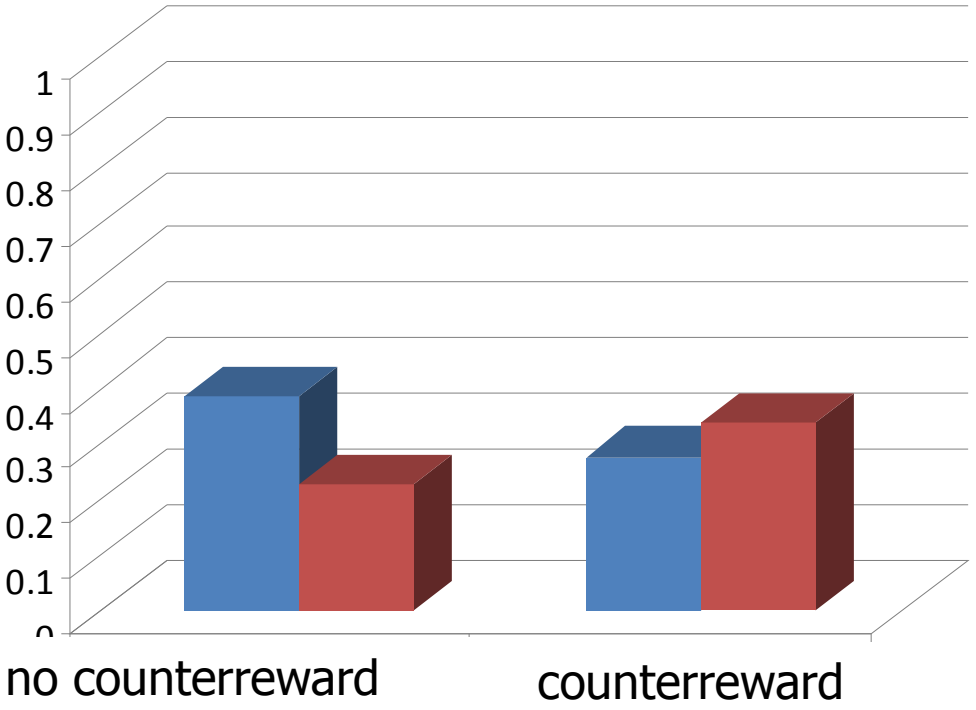
my contribution → your reward → my contribution → …

Counter-reward hypothesis
Peer reward institution is less effective if counter-reward is possible

Formal exchange models / experiments:

Flache & Macy 1996 Journal Mathematical Sociology (reinforcement learning)
Flache 2002 Journal Mathematical Sociology (strategic rationality in repeated-game)
Flache 1996; Flache & Bakker 2012 (experiment)

Counter-reward: some results from earlier experiments (Flache, 1996)



- average contribution
- average reward

Rounds 3..27

of 30 rounds game

N=70 subjects (35/35)

7 groups per condition

Flache, A. **(1996).** *The Double Edge of Networks.* Amsterdam: Thesis Publishers.

Available upon request from the author (sold out)

Is this different for counter-punishment?

Earlier counter-punishment experiments (Nikiforakis etc)

› No (truly) ongoing exchange (relabeling, focus on counter-p)

With ongoing exchange:

› Punisher faces future interactions with target, BUT …

› … so does the counter-punisher

⇒ *Punishers are strongly emotionally motivated in the first place*

Counter-punishment hypothesis
Peer punishment institution is robust even if counter-punishment is possible

Manipulating counter-sanctions in repeated game: Anonymous vs non-anonymous peer sanctioning

Anonymous (no counter sanctioning):

⇒Only past *contribution* visible

⇒Sanction can not be linked to sanction

Non-anonymous (counter sanctioning is possible):

⇒Both past *contribution* and past *sanction are* visible

⇒Sanction can be linked to *both* contribution *and* sanction

Experiment: repeated game (reward)

Per round 5 decisions per player:

Phase 1: *I contribute (yes / no)*

Phase 2:

I reward buddy 1 (yes / no)

I reward buddy 2 (yes / no)

I reward buddy 3 (yes / no)

I reward buddy 4 (yes / no)

Contribution outcome:

Number buddies who contribute	0	1	2	3	4
I do not contribute	0	6	12	18	24
I contribute	-14	-8	-2	4	10

Experiment: repeated game (reward)

Per round 5 decisions per player:

Phase 1: *I contribute (yes / no)*

Phase 2:

I reward buddy 1 (yes / no)

I reward buddy 2 (yes / no)

I reward buddy 3 (yes / no)

I reward buddy 4 (yes / no)

Contribution outcome:

Number buddies who contrib.	0	1	2	3	4
I do not contr	0	6	12	18	24
I contr	-14	-8	-2	4	10

Sanction outcome (reward):

Number of peers you reward: / *Number of peers who reward you:*	0	1	2	3	4
0	**0**	**10**	**20**	**30**	**40**
1	**-3**	**7**	**17**	**27**	**37**
2	**-6**	**4**	**14**	**24**	**34**
3	**-9**	**1**	**11**	**21**	**31**
4	**-12**	**-2**	**8**	**18**	**28**

Experiment: repeated game (punishment)

Per round 5 decisions per player:

Phase 1: *I contribute (yes / no)*

Phase 2:

I sanction buddy 1 (yes / no)

I sanction buddy 2 (yes / no)

I sanction buddy 3 (yes / no)

I sanction buddy 4 (yes / no)

Contribution outcome:

Number buddies who contrib.	0	1	2	3	4
I do not contr	0	6	12	18	24
I contr	-14	-8	-2	4	10

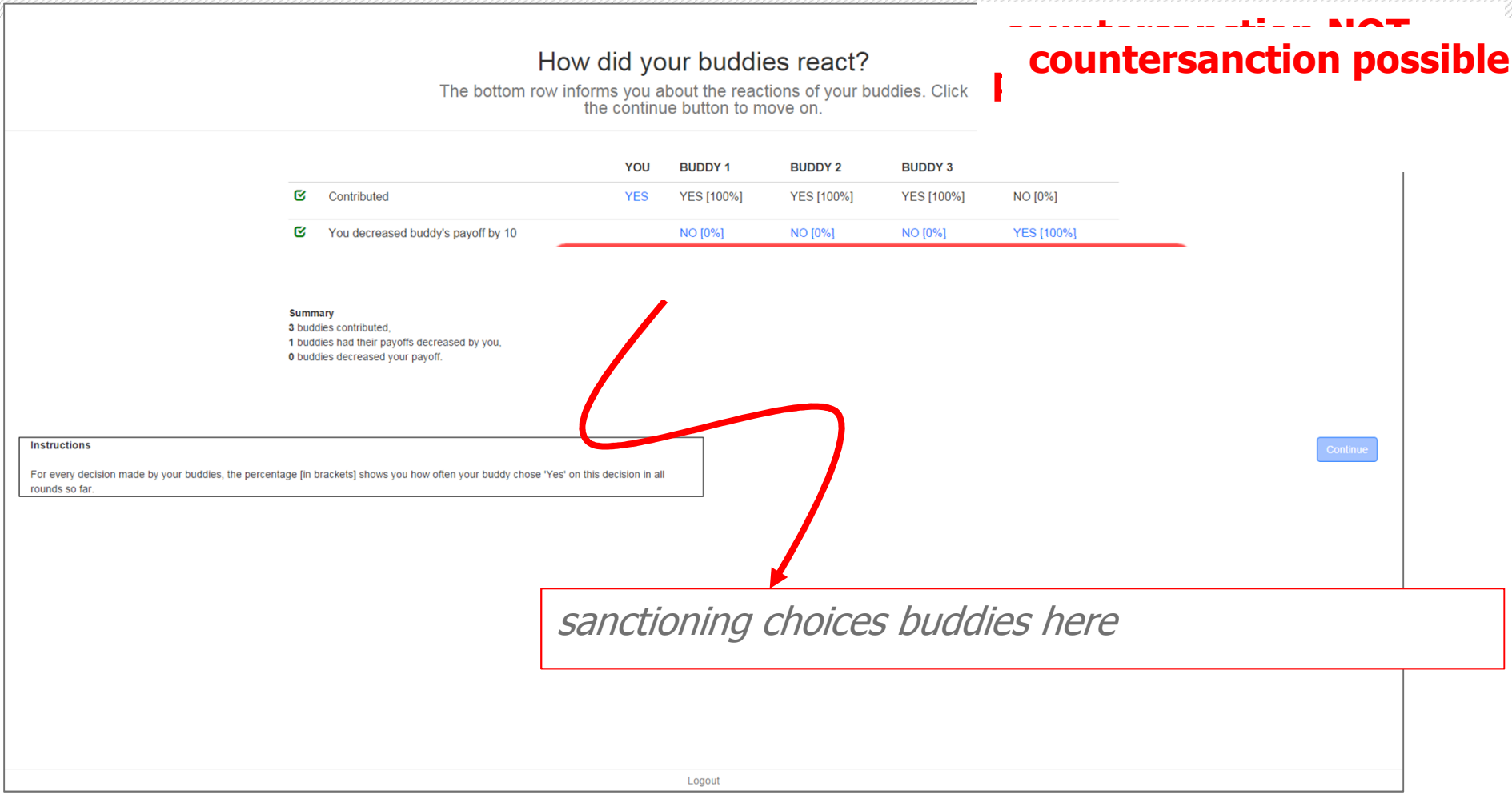
Sanction outcome (punish):

Number of buddies you punish: / *Number buddies who punish you:*	0	1	2	3	4
0	**0**	**-10**	**-20**	**-30**	**-40**
1	**-3**	**-13**	**-23**	**-33**	**-43**
2	**-6**	**-16**	**-26**	**-36**	**-46**
3	**-9**	**-19**	**-29**	**-39**	**-49**
4	**-12**	**-22**	**-32**	**-42**	**-52**

Main experimental manipulation: show sanctioning choices of “buddies”

countersanction NOT

countersanction possible



sanctioning choices buddies here

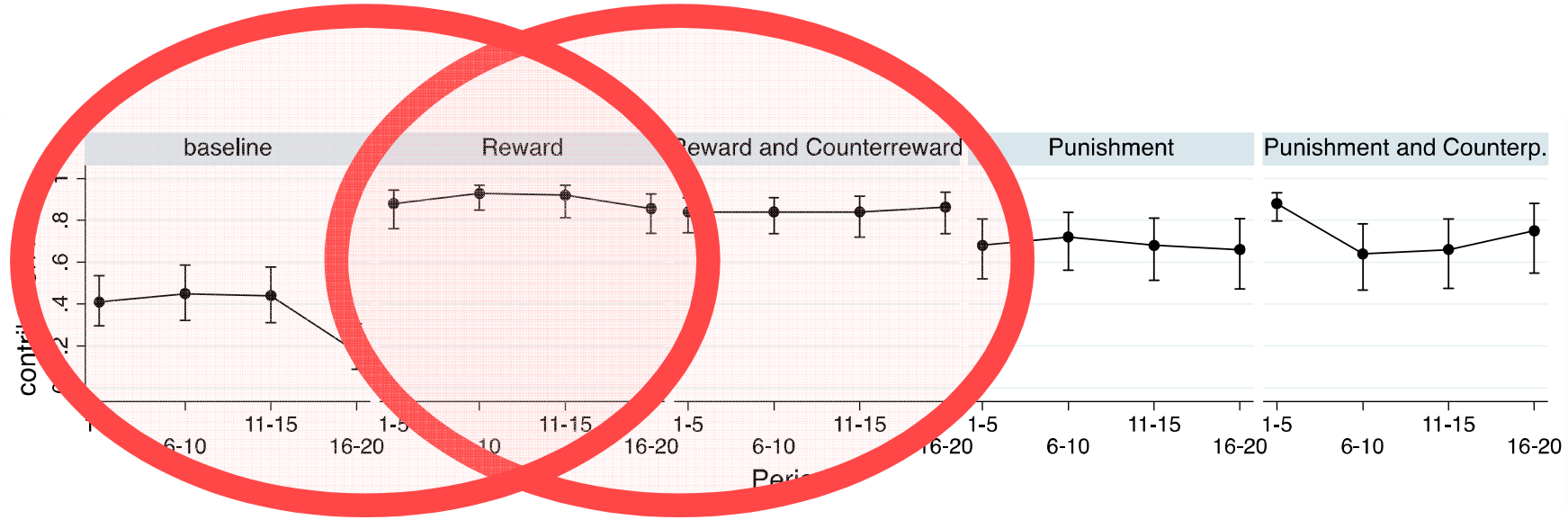
Design, Sample and Methods

Baseline	Punishment	Punishment + Counter	Reward	Reward + Counter
20	20	20	25	35

Total: 120 participants across 9 sessions

Methods of Analysis:
multilevel logistic regressions / poisson regressions (random intercepts subjects and groups)

Testing effects of (counter-)reward: reward institution is effective



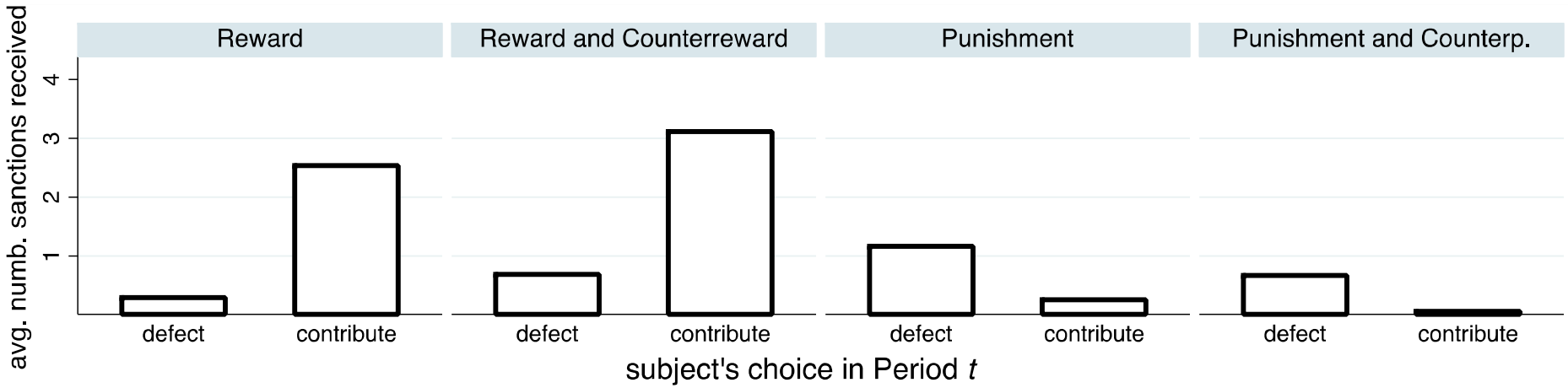
Does counter-reward reduce contributions?

Results do not support counter-reward hypothesis.

Testing peer sanction mechanism:

Do contributors receive more rewards (fewer punishments) in subsequent sanctioning stage?

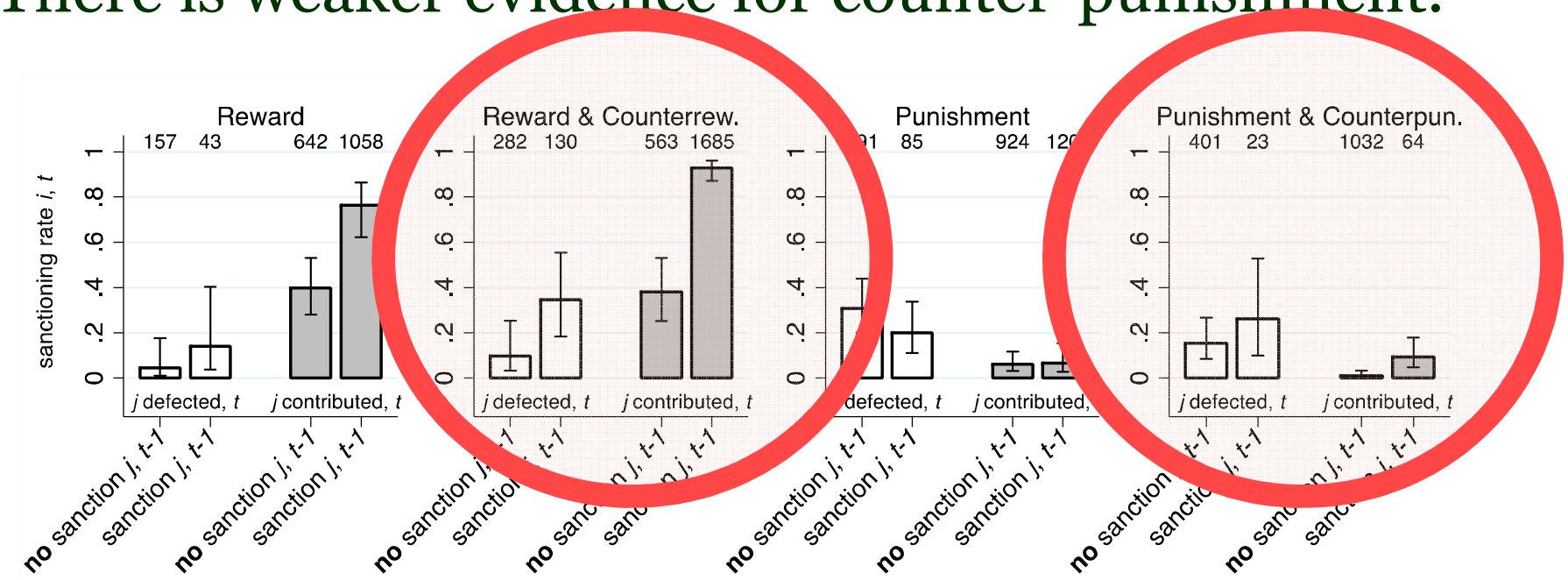
YES.



Testing counter-sanction mechanism:

Subjects rewarded by a team mate *j* in the previous period rewarded that team mate at a higher rate in the current period. **Counter-reward confirmed.**

…There is weaker evidence for counter-punishment.



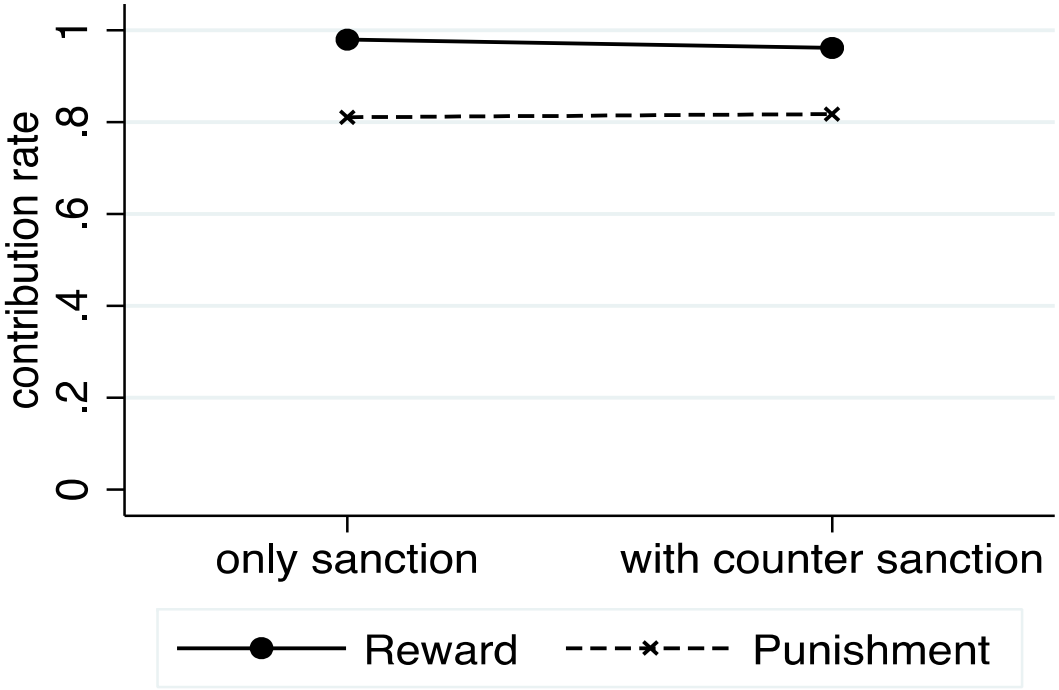
Counter-reward vs counter-punishment on contribution:

Non-anonymity will reduce rates of contribution less for a punishment institution than for a reward institution.

No support.

Peer reward not robust?
⇒ **disconfirmed**

Peer punishment robust?
⇒ **confirmed**



Take home …

- **Is the problem of counter-sanctioning overrated?**
  - Many collective action problems are "ongoing exchanges"
  - We found **no supportive evidence of negative effects of counter-sanctioning** on contributions in ongoing exchange.
  - Neither for reward (unexpected) nor for punishment
  - BUT: there was evidence for underlying mechanism
- But our results may not generalize …
  - more costly sanctions?
  - games with stronger endgame effects?
  - games where both punishment and reward are available?

Some general conclusions about experiments

- Experiments can help to uncover basic mechanisms at microlevel of complex social dynamics “building blocks”
- Test theoretically predicted effects of structural conditions: identify causal factors rather than inferring them from (lack of) correlation

Critique:

- lack of external validity
- focus on only aspect at a time, reality is more complex

⇒ My answer: only if we get the building blocks right, we can build solid greater buildings.