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

ICS
RUG / UU / RU / UvA

When contact backfires, and when it does not A social influence model of the dynamics of affective polarization

Andreas Flache* @

ETH workshop
Measuring, Modelling and Mitigating Opinion Polarization and Political Cleavage, Sep 13-15, ETH Zürich

*Joint work with Yade Rotte and Alla Loseva

Affective polarization and “spin-out”

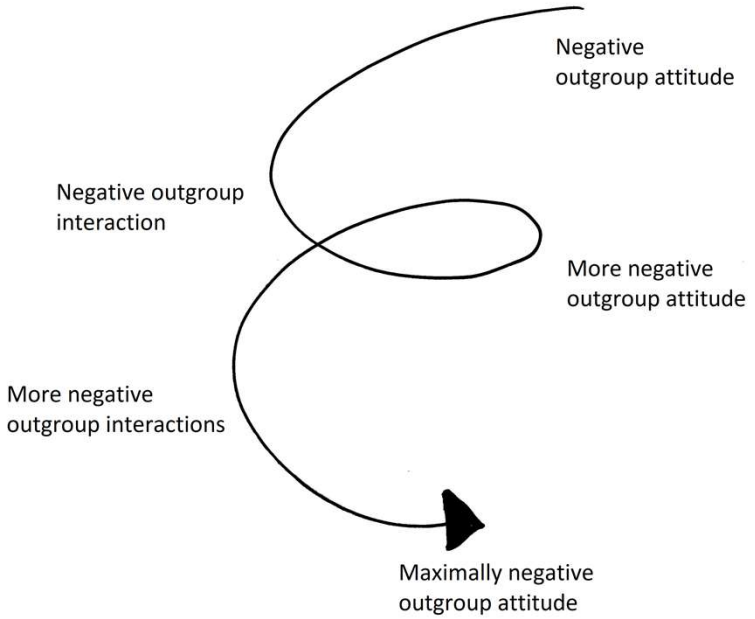
2008, Fiorina et al. 2008). But regardless of how divided Americans may be on the issues, a new type of division has emerged in the mass public in recent years: Ordinary Americans increasingly dislike and distrust those from the other party.

Democrats and Republicans both say that the other party's members are hypocritical, selfish, and closed-minded, and they are unwilling to socialize across party lines, or even to partner with opponents in a variety of other activities. This phenomenon of animosity between the parties is known as affective polarization.

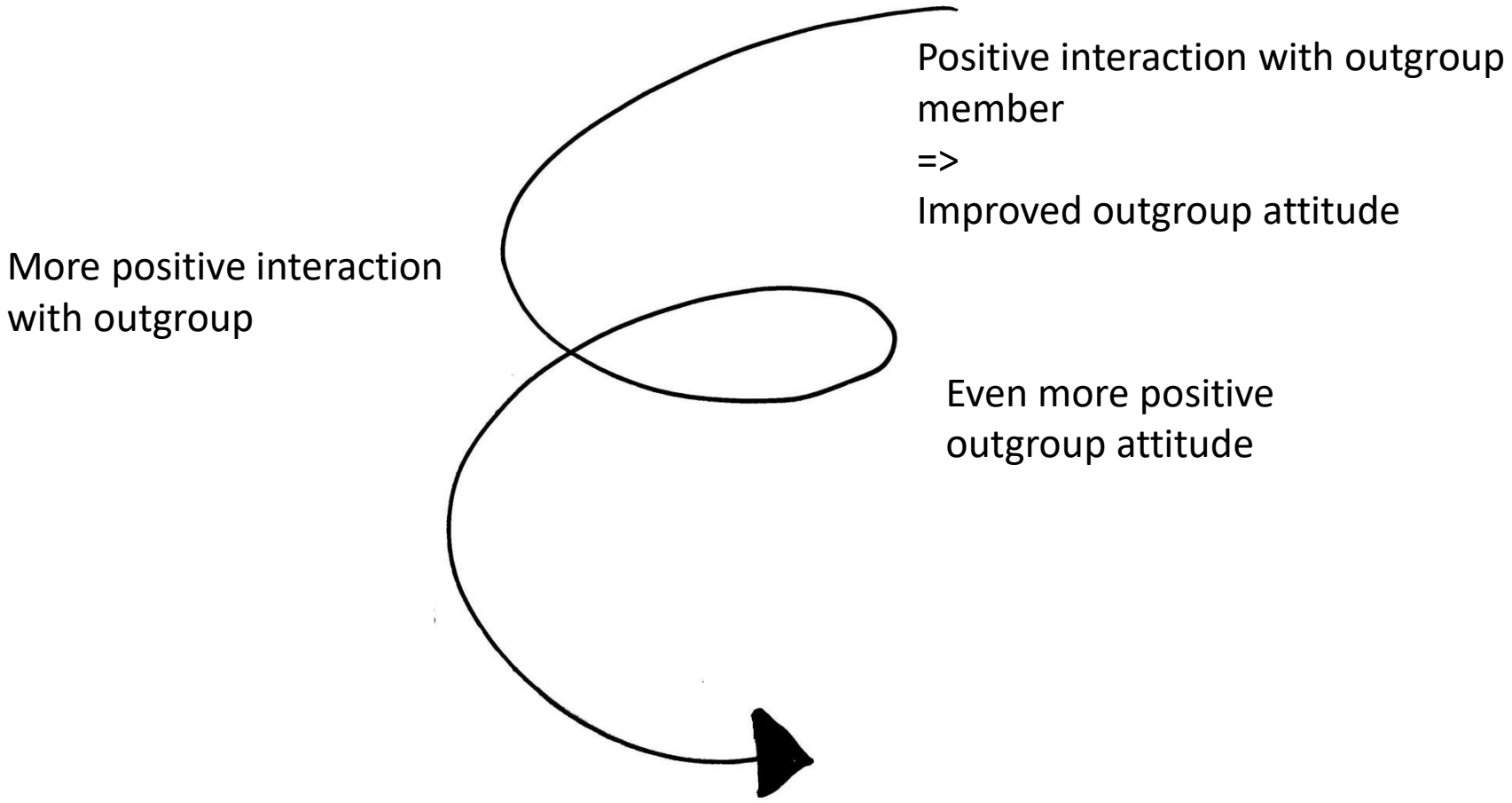
Iyengar 2019, Annu. Rev. Political Sci. 2019. 22:129–46

“Spin-out”

- A spiral of mutually reinforcing negative outgroup attitudes
- Polarisation between ethnic groups concerning the attitudes towards ethnic groups
  - Positive opinion towards the own group
  - Negative opinion towards the ethnic outgroup



The remedy? Contact theory



But then: it's more complicated

- Not only intergroup contact, but also "intragroup contact" aka socialization
- Peer influence in attitude formation
  - Adjusting to attitudes of friends
- Both intragroup attitude (towards own group) and intergroup attitude are influenced
- Social selection / homophily
  - Preferring interaction with similar others (in terms of attitudes, interests, group membership / identity)
  - Generally preferring interaction with groups one likes better
- Thus: the very network relations in which attitudes are influenced can change over time, driven by these attitudes

Aim:

- build formal theoretical model capturing the interplay of these processes
- explore theoretically conditions under which then "contact works", or … "backfires"

What happens when?
Towards an "opinion dynamics" model

Intergroup attitudes

Process and outcome of opinion dynamics

Intergroup attitudes

Process and outcome of opinion dynamics

We have models for this ***I need a model for that***

Core assumption: social influence

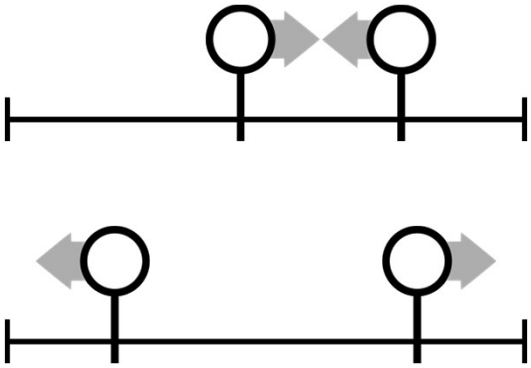
- **Positive influence on intergroup attitude**

  Small discrepancy ⇒ attraction / assimilation

- **Negative influence on intergroup attitude**

  Large discrepancy ⇒ repulsion / distancing

e.g. Macy ea 2003; Jager & Amblard, Flache & Mäs 2008, Flache & Macy 2011, Feliciani ea 2017, …



Discrepancy depends on:

⇒ opinion disagreement (e.g. about the attitude towards group X)

⇒ whether same group or not ("structural xenophobia")

⇒ **attitude towards group to which "the other" belongs:**

→and this is also one of the opinions that is influenced

Advances in Complex Systems
Vol. 21, Nos. 6 & 7 (2018) 1850017 (32 pages)

DOI: 10.1142/S0219525918500170

World Scientific
www.worldscientific.com

ABOUT RENEGADES AND OUTGROUP HATERS: MODELING THE LINK BETWEEN SOCIAL INFLUENCE AND INTERGROUP ATTITUDES

ANDREAS FLACHE

Department of Sociology,
University of Groningen, Grote Kruisstraat 2/1,
9712 TG Groningen, The Netherlands
a.flache@rug.nl

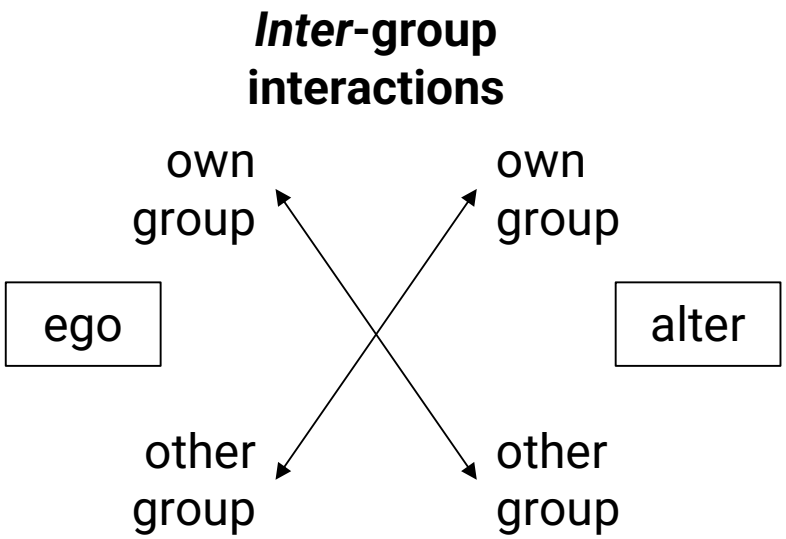
Somewhat older working paper version:

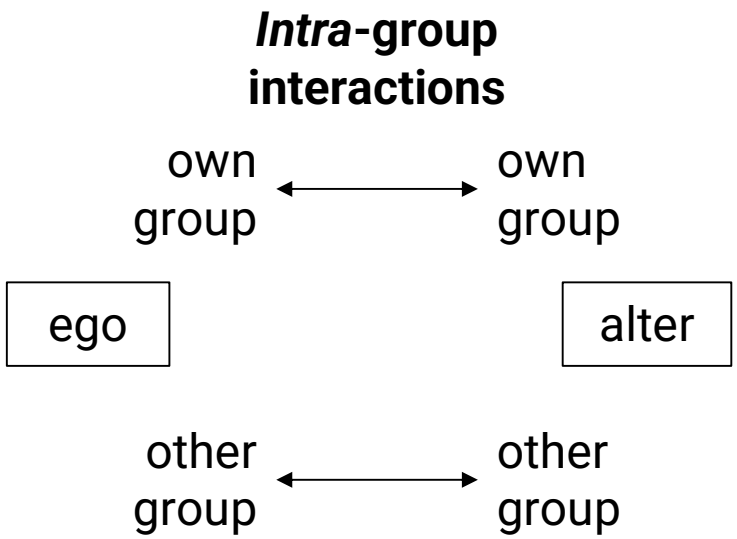
arXiv.org > physics > arXiv:1708.03917

A more realistic version for this paper:

- Everyone has 2 attitudes: in- and outgroup
- Both attitudes are subject to social influence
- Both attitudes affect how open we are towards being influenced by “the other” (in-in, out-out).
- Social selection: agents choose interaction partners based on preferences (i.a. intergroup attitudes)

A social influence model of inter- and intra-group “contact”





Intergroup attitudes have "social impact"

On influence:

- Positive attitude about your group: positive influence more likely
- Negative attitude about your group: negative influence more likely

On social selection:

- Positive attitude about your group: interaction more likely
- Negative attitude about your group: interaction less likely

Questions:

Does contact still “work” if:

- Social impact of improved outgroup attitudes is limited by “structural xenophobia”?
- People can not always select whom they want to interact with (e.g. segregation)
- Not everyone likes their own group (“ingroup critics”)

Model ingredients (1)

Figure 3. Schematic representation simulation algorithm:

1. Assign group membership *g* and initial att_0 and att_1 to every agent based on population composition and Beta distributions.

Repeat until simulation stops:

Two groups only!

2. Pick an agent *i* at random
3. Agent *i* selects one interaction partner *j* from entire population

3.1 compute for every other agent *k* in the population the attractiveness as potential interaction partner for *i*, based on Equations 5 and 6

3.2 Select 1 out of the *N*-1 others as interaction partner *j*. The more attractive *j* is for *i*, the more likely *j* will be picked. Probabilities are calculated based on Equation 7.

4. Interaction and influence *i*-*j*:

4.1. For both *i* and *j*: Compute discrepancies d_{ij}, d_{ji} with the other agent ($0 <= d <= 1$) as given in Equations 3 and 4.

4.2. Social influence: adapt for both agents att_0 and att_1 as given in Equations 1 and 2.

Go back to "repeat until … "

Model ingredients (2): initial intergroup bias

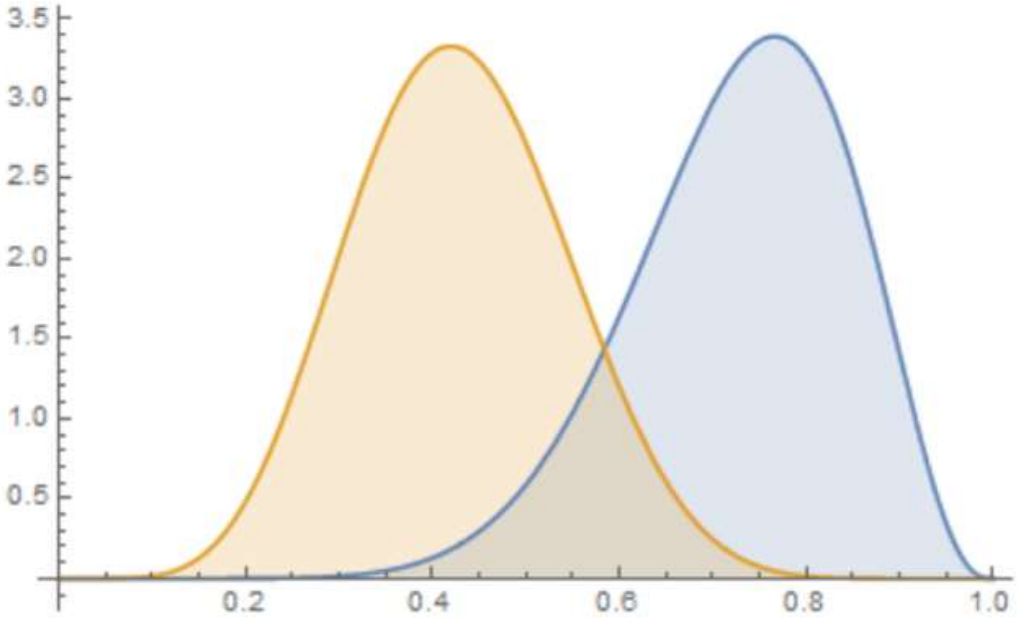


Figure 2. Distribution initial attitudes in baseline scenario of simulation experiments. Blue: Ingroup attitudes, Beta(10, 3.75). Orange: Outgroup attitudes, Beta(7.5,10).

Modelling subjective discrepancy

"Raw" discrepancy *i* towards *j*:

$$d_{ijt} = \beta_A \left| g_j(1 - att_{1it}) + (1 - g_j)(1 - att_{0it}) \right| + \beta_D \left| (g_j - g_i) \right| + \beta_O \, dis_{ijt}$$

attitude towards group of "other"

$$\beta_O + \beta_D + \beta_A = 1, \quad \beta > 0.$$

dis_{ij} = *average disagreement i-j accross both attitudes*

β_O = impact of opinion disagreement on discrepancy

β_D = "fixed xenophobia": impact of "same group" (g_i , $g_j \in \{0,1\}$) on discrepancy

β_A = social impact of intergroup attitude towards group of *j* on discrepancy

Discrepancy higher if:

- **We disagree more on both groups** (β_O) individual disagreement
- **We are not same group** (β_D) structural xenophobia
- **I like your group less** (β_A) social impact intergroup attitude

Model ingredients (4): influence

Low discrepancy: attitudes of i and j become more similar

High discrepancy: attitudes become more dissimilar

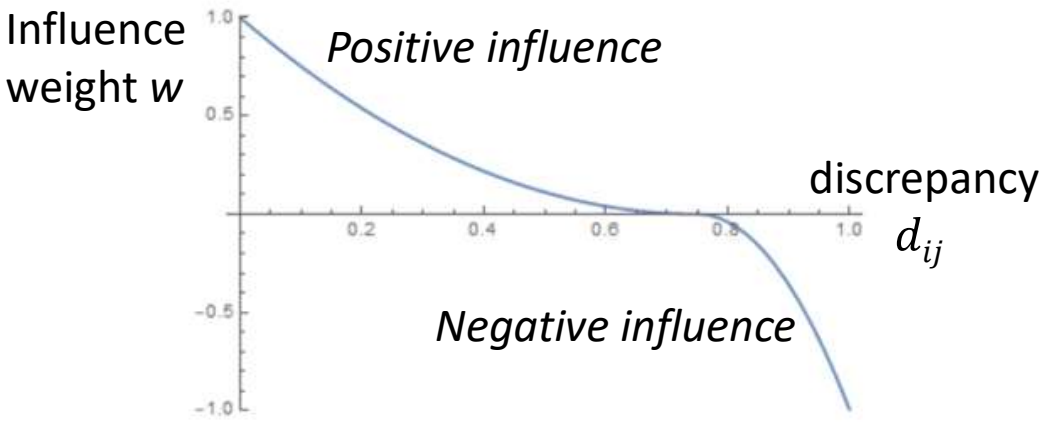


Figure 4: influence-weight function f(d) with *s*=2 and *t*=0.75

"Psychological realism":

- Ingroup attitudes change slower than outgroup attitudes
- Negative influence only if discrepancy is really high

Model ingredients (5): selection

The lower the discrepancy i-j, the more “attractive” j is as an interaction partners for i.

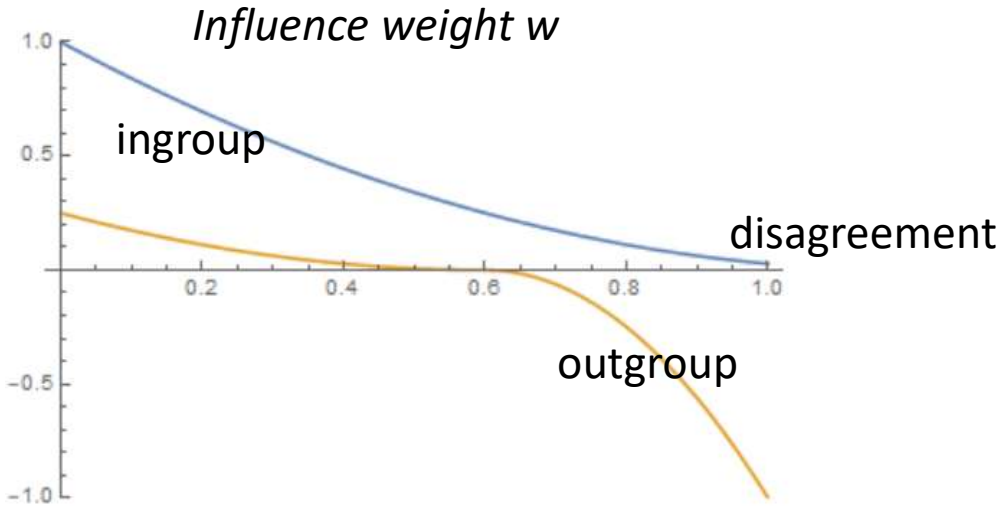
Ego (i) selects 1 Alter (j) for interaction. The more attractive, the more likely.

Model parameter “hs” scales how much impact preference has on selection decision

Baseline scenario

- 2 groups
- N = 110 (55/55)
- Mild “fixed xenophobia” (betaD = 0.375)
- Intergroup attitudes have no direct social impact (betaA = 0)

Can contact still work?



Baseline scenario: contact works

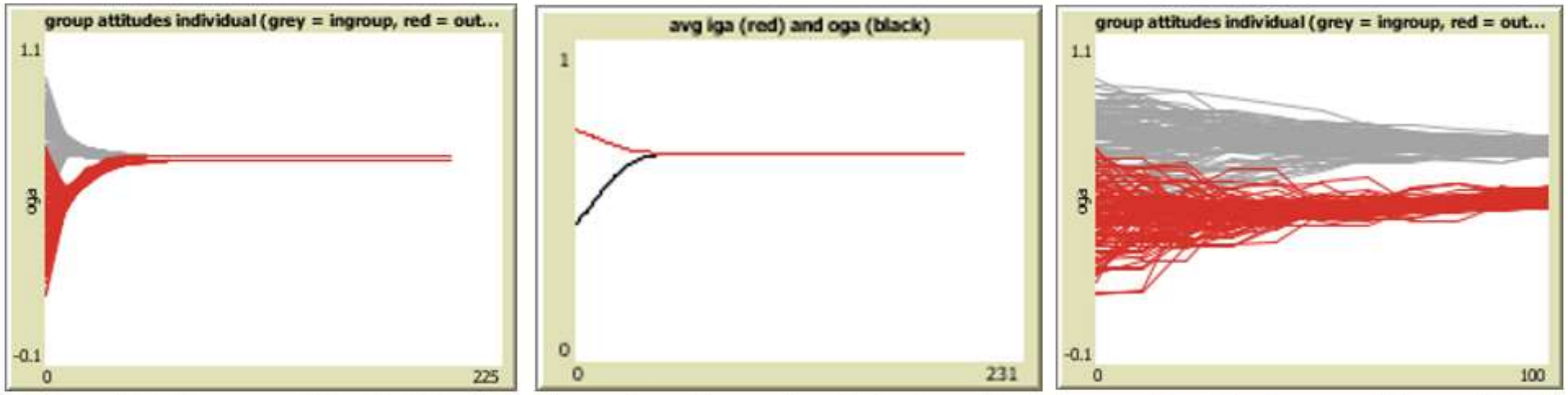


Figure 5. Change of individual iga (grey) and oga (red) in baseline scenario, single run ($\beta_D = 0.375$). Left: single trajectories, 20.000 simulation events, Middle: change of average iga and oga, Right: single trajectories, first 1000 simulation events.
Notice that every individual is represented twice here: with an ingroup attitude and an outgroup attitude. Unit x-axis: 100 simulation events. (** new basline etc **)

But what happens if structural xenophobia becomes stronger? (Increase betaD, all other things equal)

How contact “backfires”

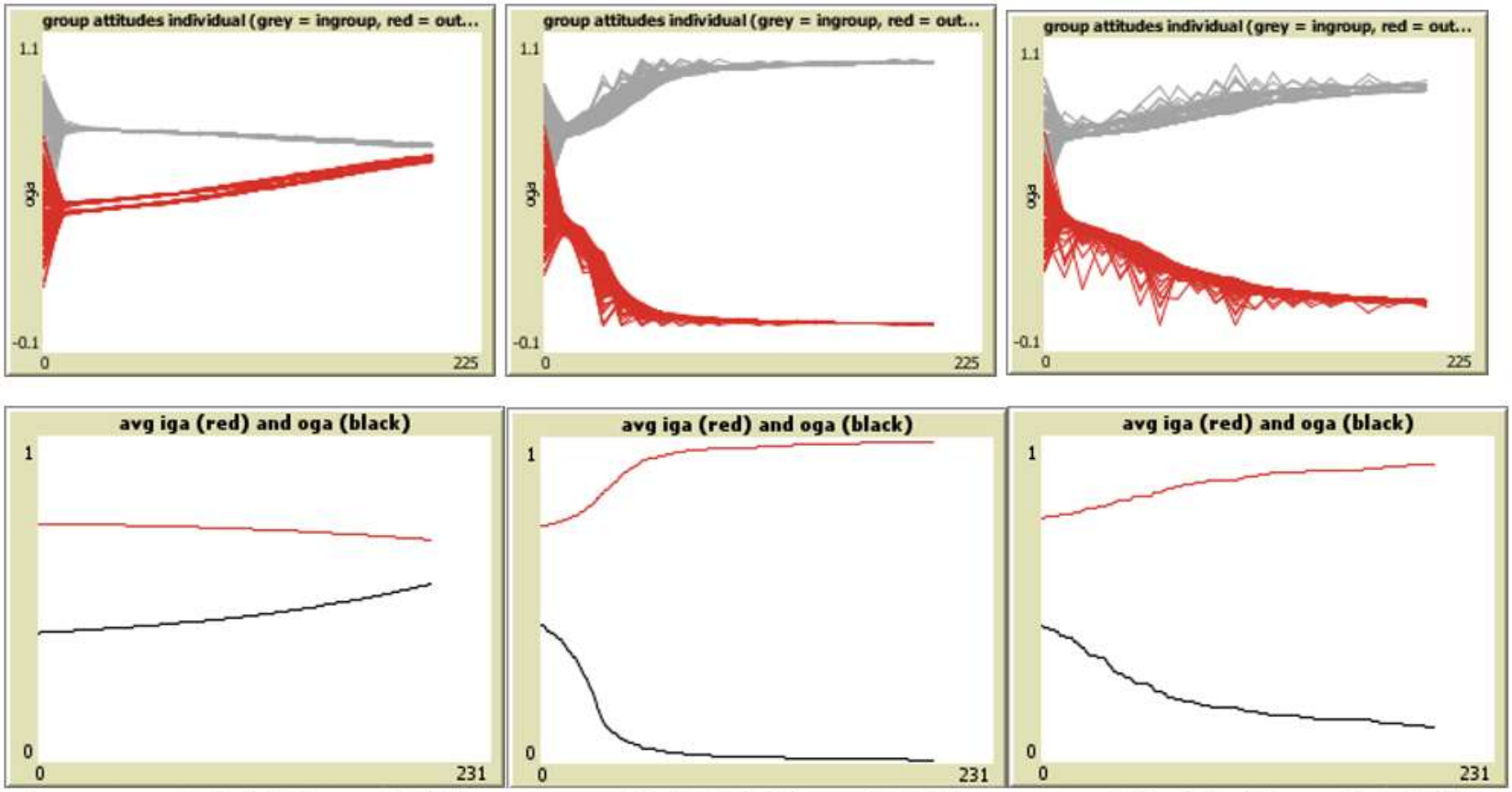
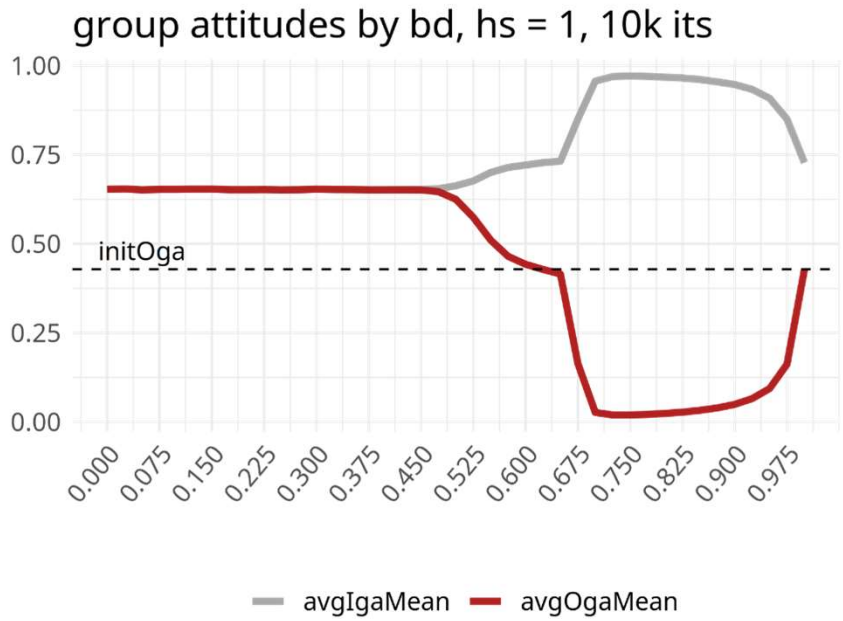


Figure 6. Left (a): ($\beta_D = 0.55$). Middle (b) ($\beta_D = 0.75$). Right (c): ($\beta_D = 0.975$).

Stronger structural xenophobia, more affective polarization? Not quite



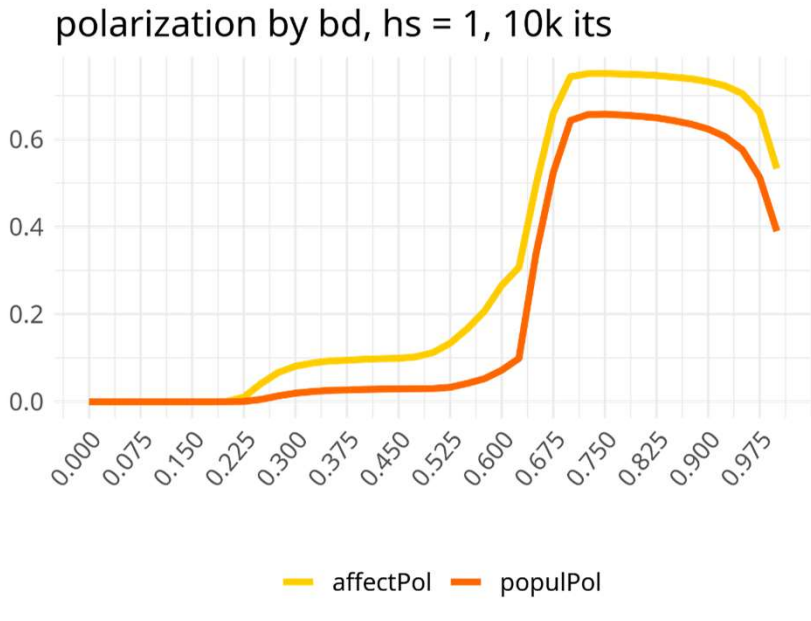
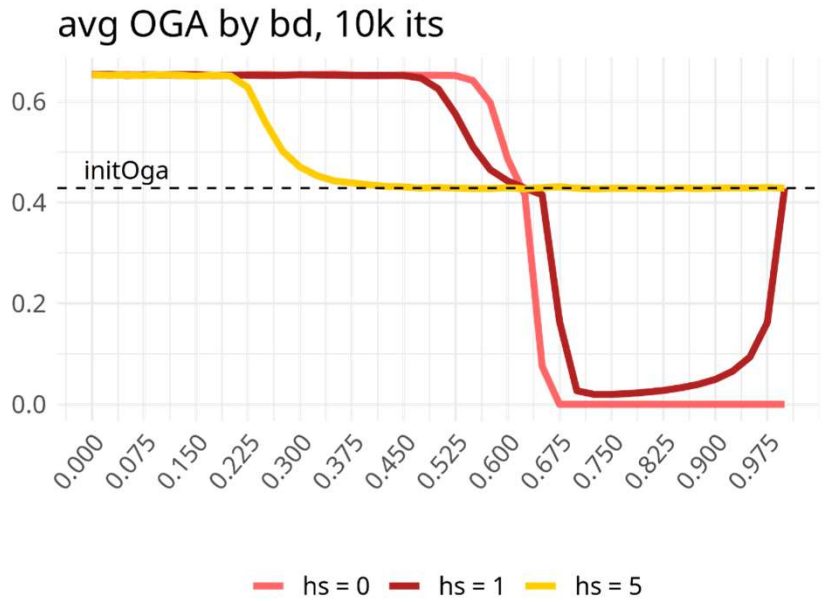
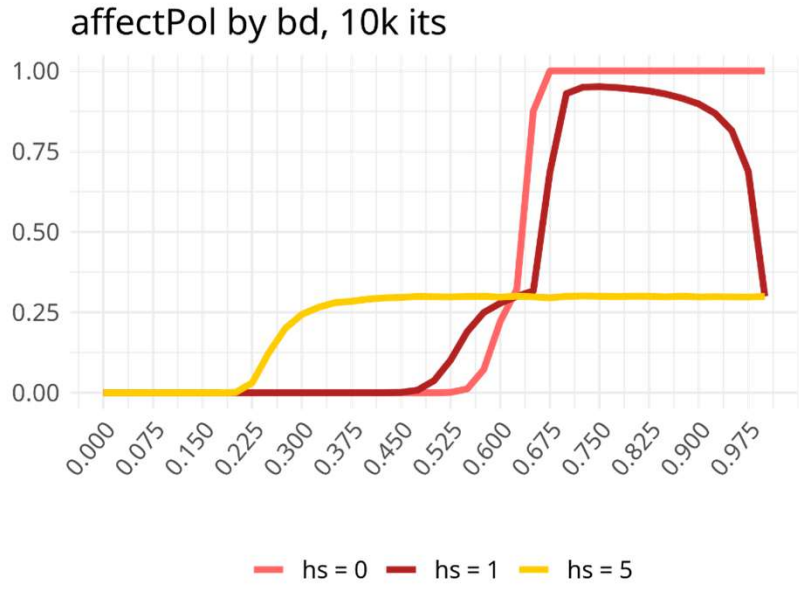


Figure 7. Experiment 1.1: effect of strength of the social impact of structural xenophobia (β_D) on intergroup attitudes (left) and polarization measures (right). Averages of 100 realizations per level of β_D after 10k simulations events.

And what if possibilities for social selection are limited?

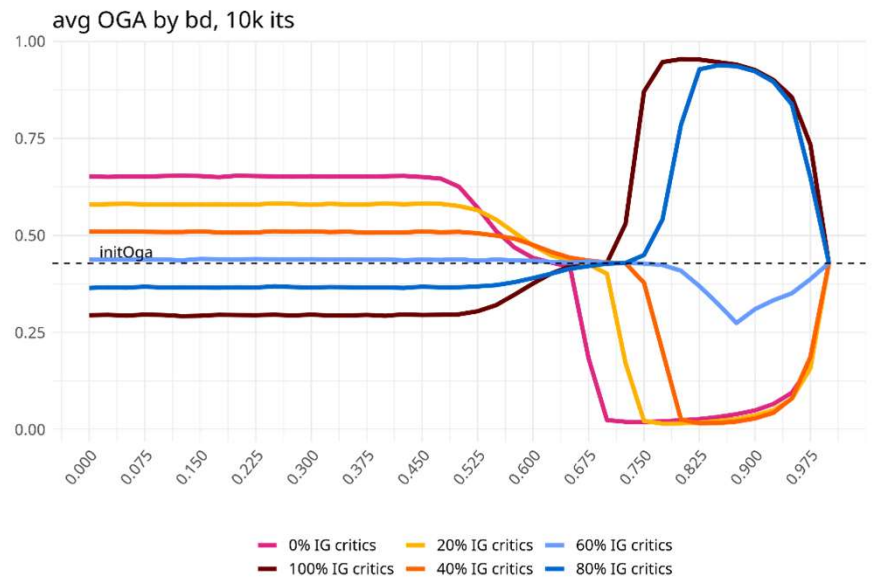


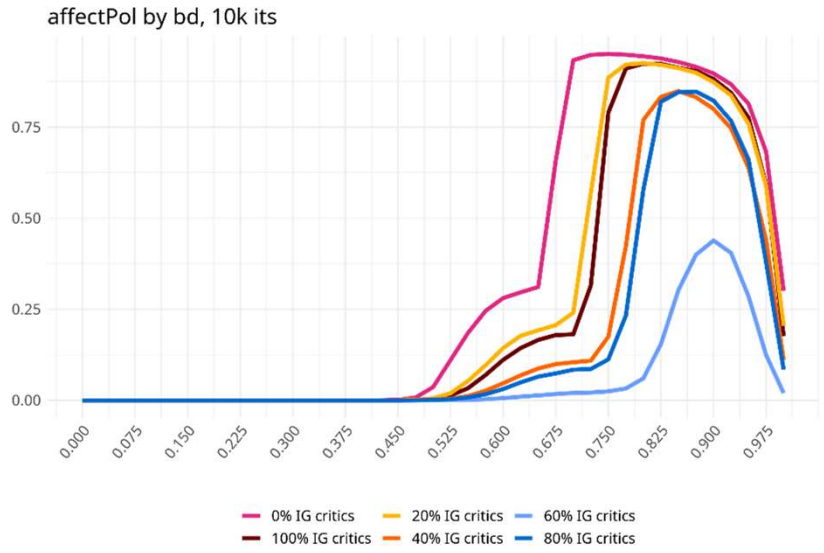


The higher hs, the more selection is based on preference (here: "outgroup avoidance")

Takeaway:
Weak structural xenophobia: outgroup avoidance undermines positive contact effect
Strong structural xenophobia: outgroup avoidance helps preventing affective polarization

And what if groups contain more “initial ingroup critics”?





Ingroup critic: initial ingroup attitude uniform random (0, 0.5) -> much less than “normal”

Takeaway:

More ingroup critics -> less positive outgroup attitudes if contact is positive (low xenophobia)

More ingroup critics -> less affective polarization if contact is negative (strong xenophobia)

And then there is “inverted affective polarization” (80%, 100%)

“Inverted affective polarization”

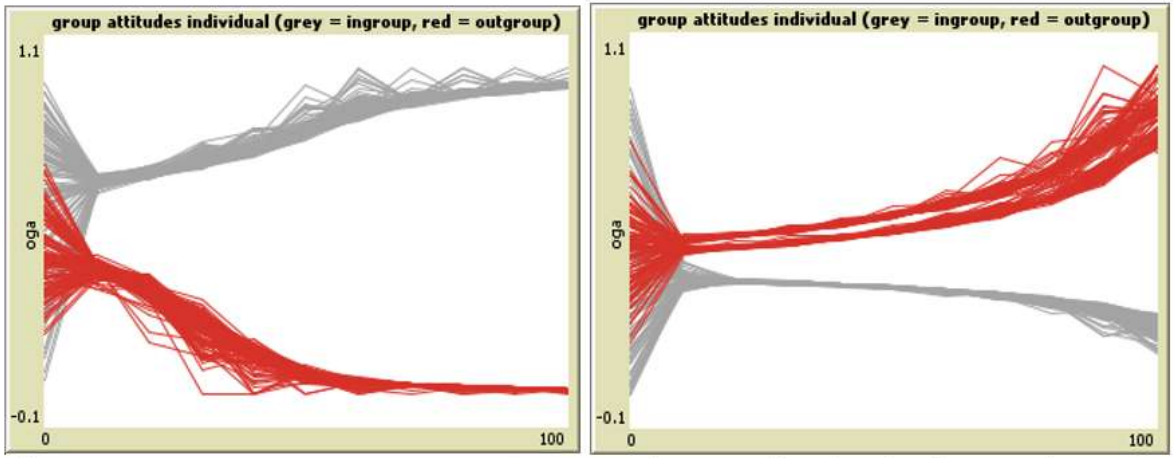


Figure 10. Representative runs showing change of individual in- and outgroup attitudes over time at betaD = 0.7 (otherwise baseline). Left: affective polarization with 0% initial ingroup-critics. Right: Inverted affective polarization with 80% initial ingroup-critics.

Everyone ends up hating the ingroup and loving the outgroup. Crazy? Maybe, but logical.

Because …

So far so good

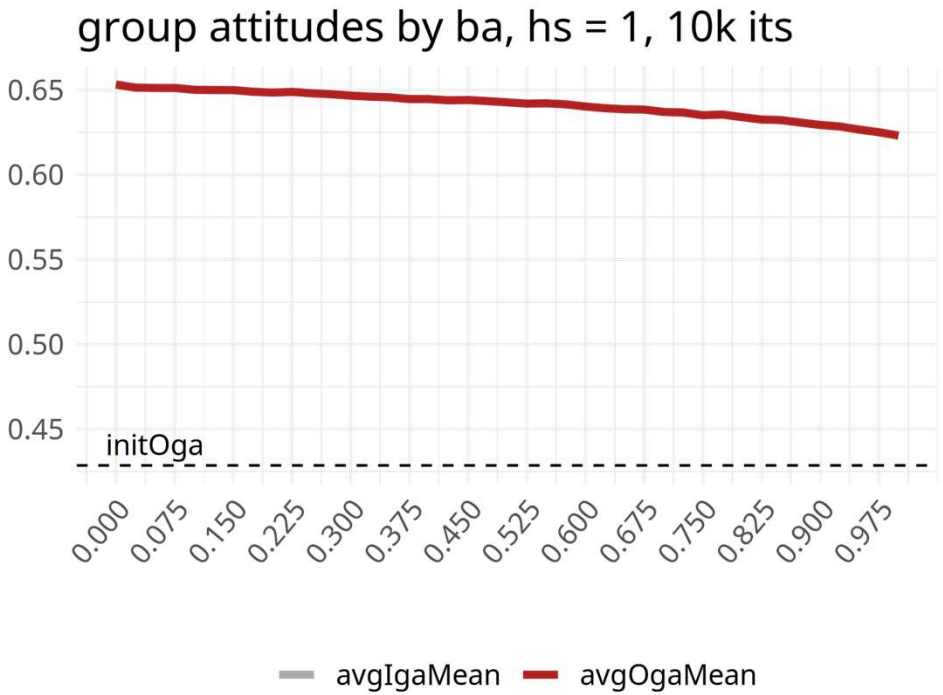
Contact can improve outgroup attitudes despite structural xenophobia (no social impact of improved oga's)

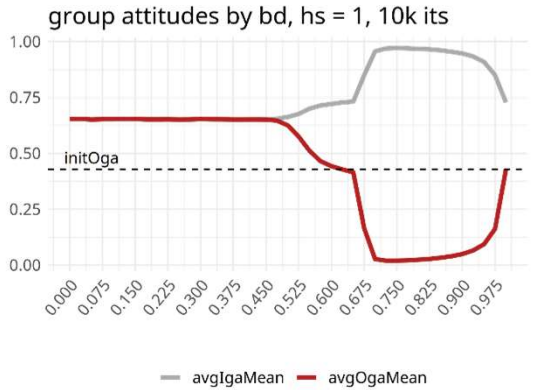
If agents can choose interaction partners, strong structural xenophobia can actually mitigate affective polarization

Ingroup critics also can mitigate affective polarization, but only when structural xenophobia is strong

But this was all about "structural xenophobia"
What if intergroup attitudes have more social impact?

Then the world becomes almost flat …

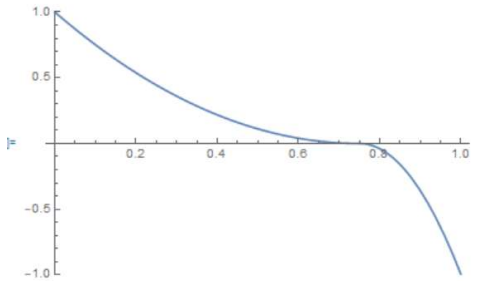




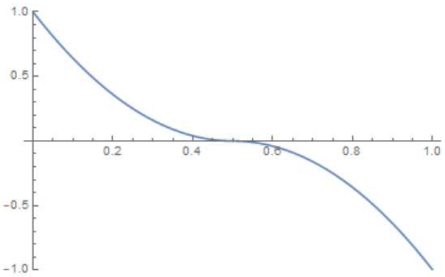
Cet-par replication experiment 1.1, but now we vary betaA 0..1 and let betaD = 0. Discrepancy now directly affected by intergroup attitude, no structural xenophobia

Takeaway: as long as intergroup attitudes are not extremely ingroup-biased, prospects for “contact works” are much better than with structural xenophobia, even when iga have extremely strong social impact (betaA = 1).

However, what if negative influence is triggered more easily?



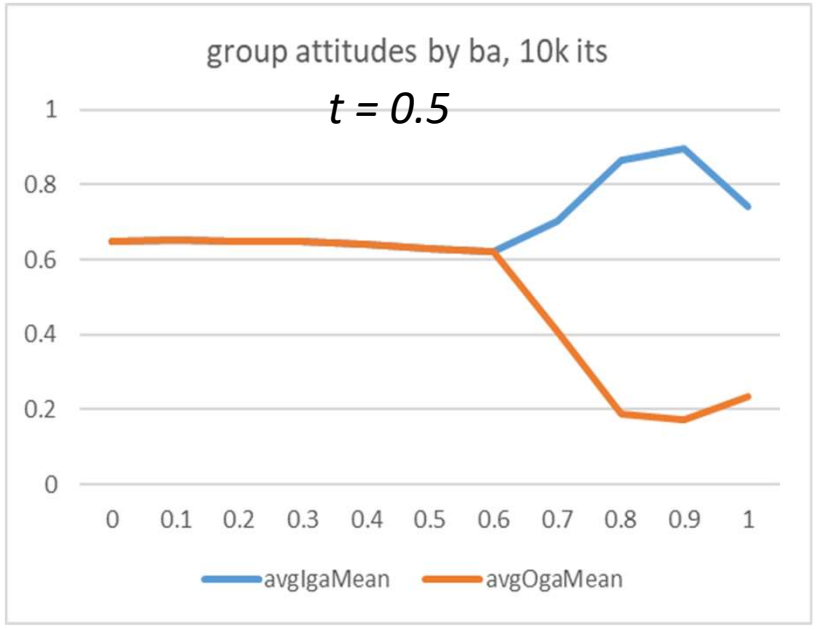
threshold t = 0.75

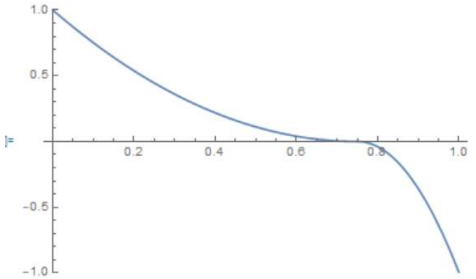


threshold t = 0.5

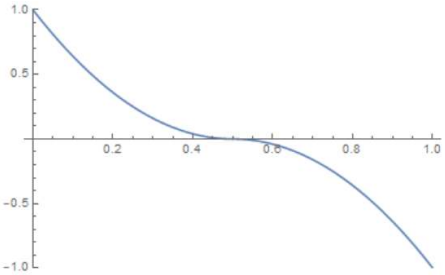
Influence weight as function of discrepancy

However, what if negative influence is triggered more easily?





threshold t = 0.75



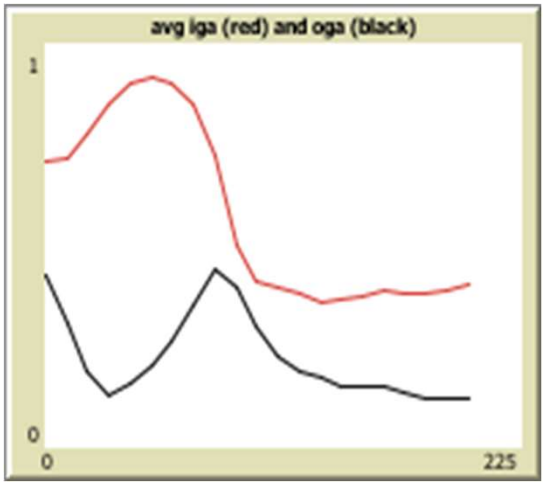
threshold t = 0.5

Influence weight as function of discrepancy

Then effects of stronger social impact are again very similar to what we found for "structural xenophobia"

Takeaway: if there is no structural xenophobia, prospects for "contact works" are better, but … only if the threshold for negative influence is very high. Otherwise: strong social impact of intergroup attitudes also leads to affective polarization.

Yet, the mechanism why very strong social impact reduces affective polarization is very different

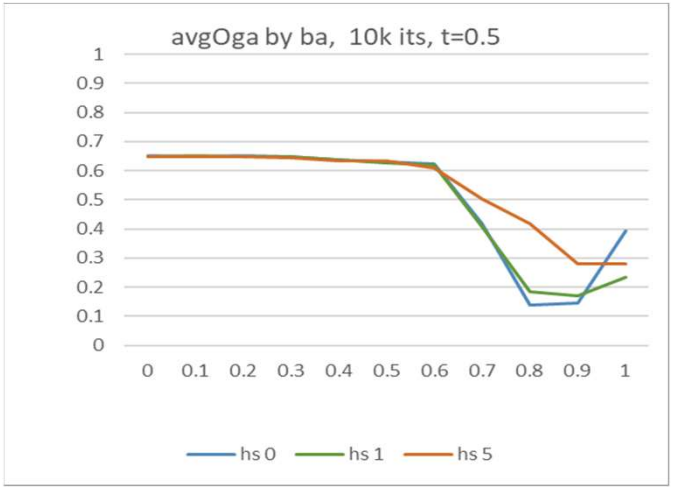


Illustrative run for betaA = 0.975, betaD = 0.

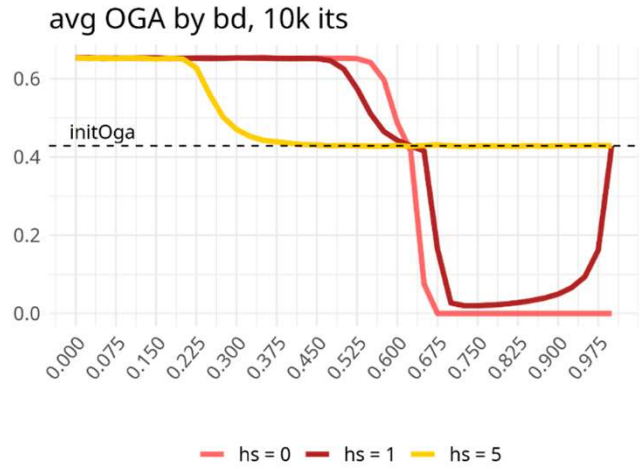
Takeaway:

- Very strong structural xenophobia suppresses interaction between groups -> less aff pol
- Very strong social impact of iga's makes groups split between "critics" and "normally biased" agents, which produces unstable dynamics, resulting in less extreme aff pol.
- "Universal critics" become possible

And what if possibilities for social selection are limited? (reloaded …)



impact iga

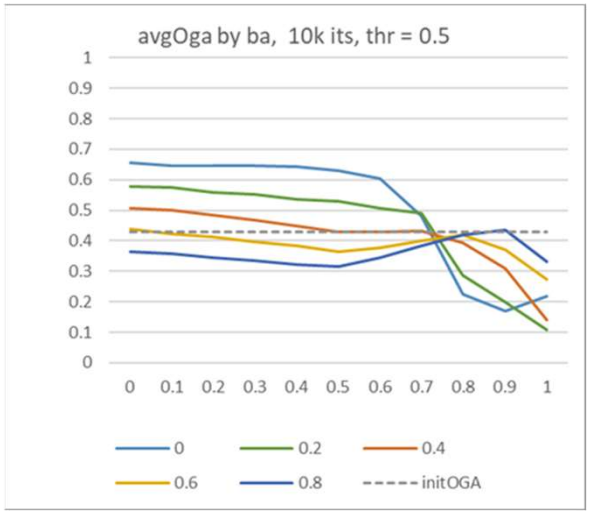


structural xenophobia (t = 0.75)

Takeaway:

- Qualitative effects look similar, but there are differences …
- Especially: this time allowing agents to enact social selection preferences does not help preventing affective polarization at high levels of betaA

And what if there are more ingroup critics? (reloaded …)



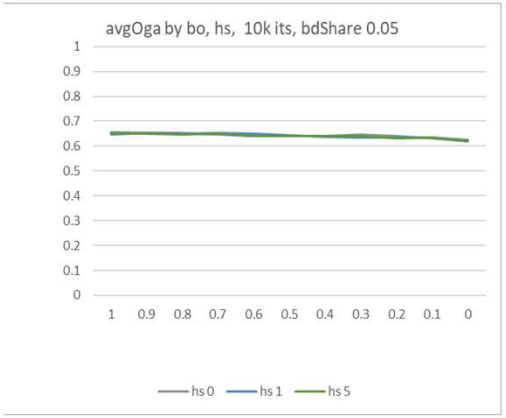
impact iga

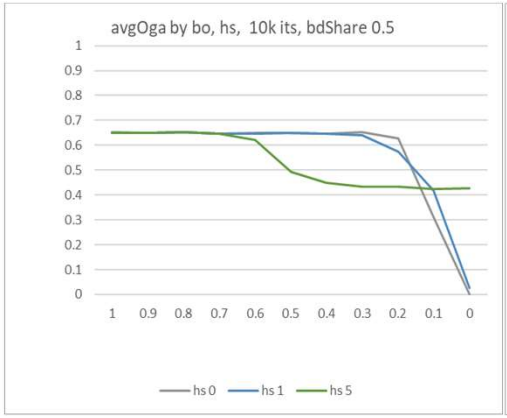
structural xenophobia (t = 0.75)

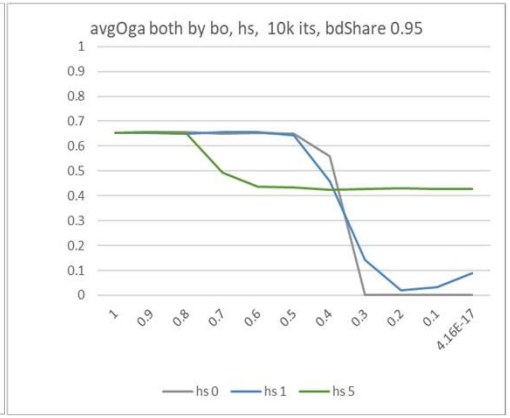
Takeaway:

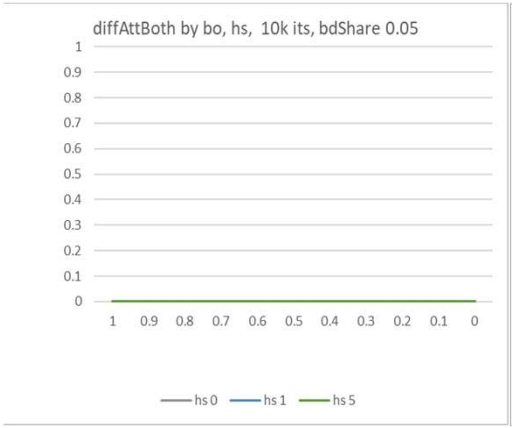
- Qualitative effects looks a bit similar, but there are differences …
- Especially: this time we do not see “inverted affective polarization”
- Reason: initially relatively mild oga’s do not create such a strong urge to distance oneself from the outgroup (as compared to strong structural xenophobia).

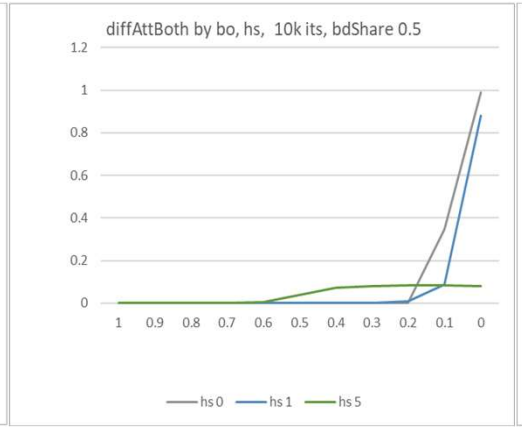
Finally: gradual shift from “attitudinal” impact to structural xenophobia (t=0.75)

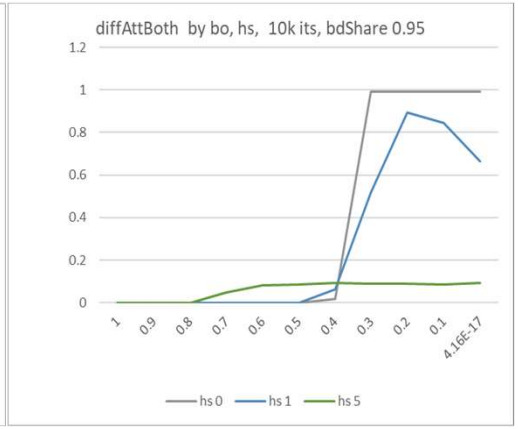






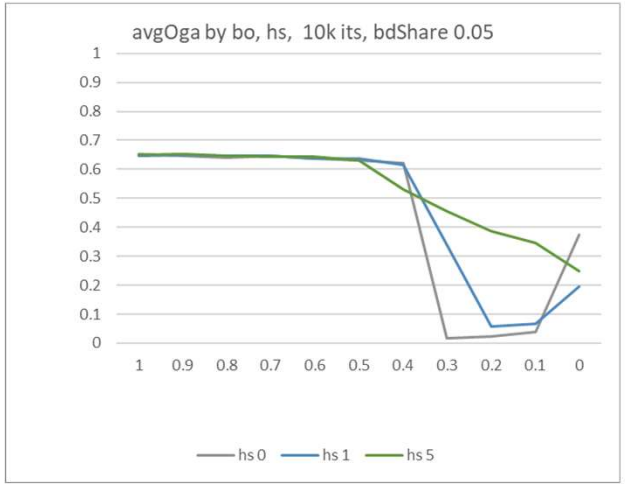


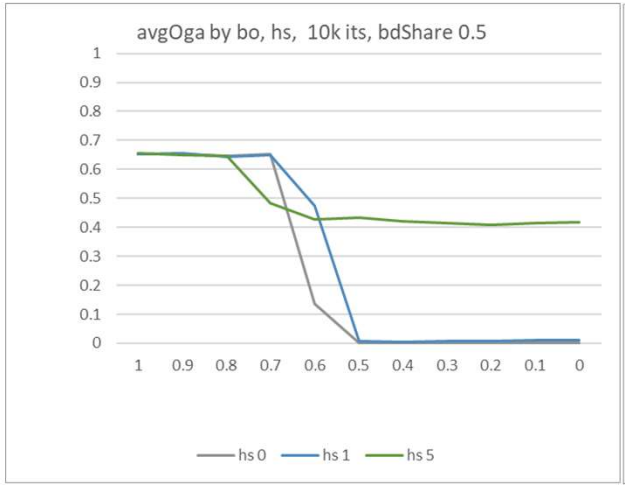


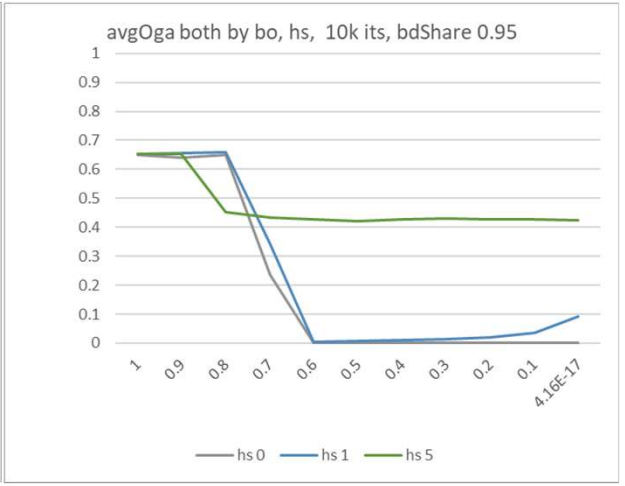


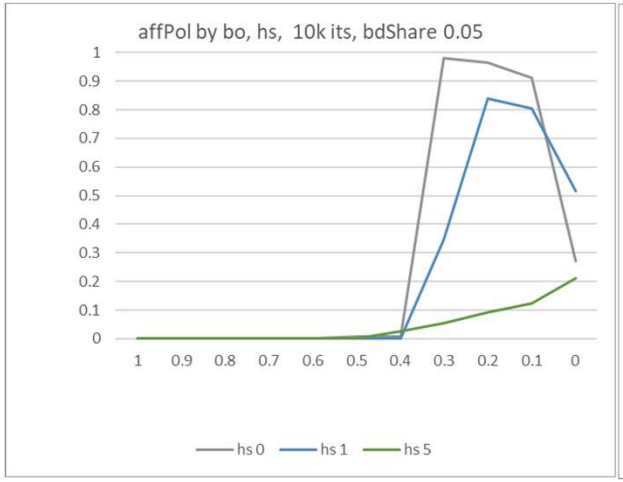
From “contact works” to “affective polarization” if threshold for negative interactions is high (t=0.75)

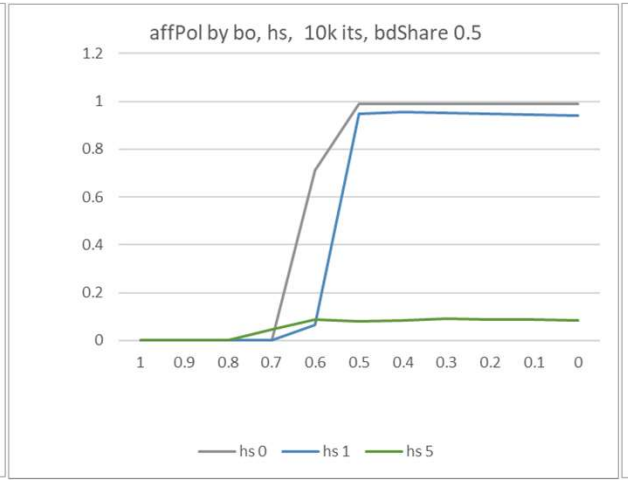
Finally: gradual shift from “attitudinal” impact to structural xenophobia (t=0.5)

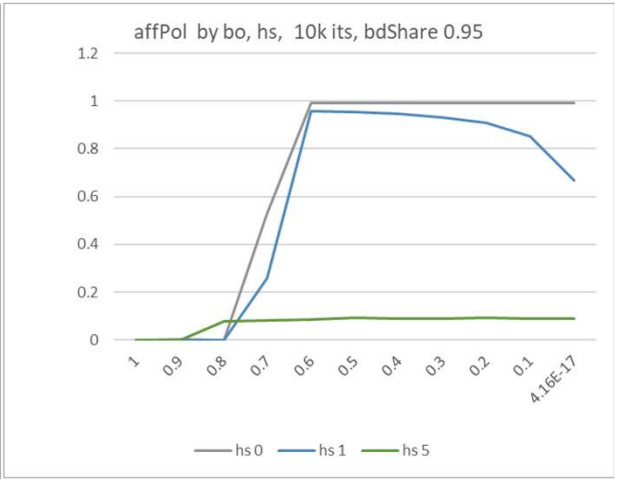












Not so much changes if threshold for negative influence is lower (t=0.5)

Conclusion and outlook

- A model integrating intra- and intergroup influence, social selection, positive and negative contact and influence
- On the whole prospects for contact to work are better when xenophobia can be “unlearned” (not structural)
- Except: if agents can choose interaction partners, strong structural xenophobia can actually mitigate affective polarization
- Funny stuf can happen: reversed affective polarization, universal critics arising.
- Of course very much is still very unrealistic
  - 2 groups
  - Influence function (negative?)
  - Agents can not “disidentify” …