

Rainer Hegselmann

9 Polarization and Radicalization in the Bounded Confidence Model: A Computer-Aided Speculation

Abstract: In the bounded confidence model agents update their opinions by averaging over all opinions that are not too far away from their own opinion. The article gives a precise definition of the basic model, offers several interpretations of the model, and introduces two simple extensions that allow to analyze polarization and radicalization. The basic model and its two extensions are seemingly simple. But the simplicity is deceptive. Lots of counterintuitive effects come as a surprise. Additionally, the article demonstrates that in terms of explanatory understanding of mechanisms, it makes a lot of sense to work with deterministic idealizations of random start distributions.

9.1 Introduction

The so-called *bounded confidence model* (BC model) spells out a simple idea: In their ongoing exchange of opinions, individuals (or *agents*) take seriously those others whose opinions are not too far removed from their own. The core of this article are two extensions of the basic BC model. The *first* extension (Section 9.2) introduces a *bias* of the type: “Leftists listen more to the left, rightists listen more to the right.” We show how, given the bias, strongly polarized camps of radicalized agents can evolve endogenously. In the *second* extension (Section 9.3) a group of radicals enters the field. Different from normal agents, the radicals simply stick to their radical opinion – no other view is taken seriously. We analyze how normal agents that take seriously opinions that are not too far removed from their own opinion, may or may not become radical.

This article also has a *methodological* objective. I want to demonstrate that the following is a fruitful approach: *First*, we initialise all our opinion dynamics with the same, constant start distribution. It is a very special, namely *representative* start distribution: it idealizes deterministically high numbers of random start distributions. *Second*, in tiny steps, a selected parameter is varied. Then, by direct comparisons of single runs of the system, we try to get an understanding of the dynamics. In contrast to laboratory experiments, such an approach is easy to implement with the help

Note: Parts of this article draw on two articles that I wrote together with Ulrich Krause, namely Hegselmann and Krause (2002, 2015). I want to thank Matthew Braham and Igor Douven for all their help.

Rainer Hegselmann, Frankfurt School of Finance & Management

of computers. Our computational analysis of polarization and radicalization in Sections 9.2 and 9.3 applies this approach. In the concluding Section 9.4 we discuss the advantages, limits, and dangers of our strategy.

Finally a warning: What follows is not the presentation of an empirically calibrated model. It is more a *computer-aided speculation* on the consequences of mechanisms that, as some empirical findings suggest, seem to be at work in the real world. I analyze the idealized mechanisms in an artificial world wherein nothing else interferes. The hope is to get thereby a better feeling for what might be at work in the real world. However, there is also bad news: My computer-aided speculations suggest that there might exist a fundamental problem for both the understanding of and intervening in polarization and radicalization processes.

9.1.1 The bounded confidence model: A formal description and some possible interpretations

Stated in a precise language, the constitutive assumptions of the BC model are:

1. There is a set I of n agents; $i, j \in I$.
2. Time is discrete; $t = 0, 1, 2, \dots$
3. Each individual starts at $t=0$ with a certain opinion, given by a real number from the unit interval; $x_i(0) \in [0, 1]$.
4. The profile of opinions at time t is $X(t) = x_1(t), \dots, x_i(t), \dots, x_j(t), \dots, x_n(t)$.
5. Each agent i takes into account only ‘reasonable’ others. Reasonable are those individuals j whose opinions are not too far away, i.e. for which $|x_i(t) - x_j(t)| \leq \epsilon$, where ϵ is the *confidence level* that determines the size of the *confidence interval*.
6. The set of all others that i takes into account at time t is:

$$I(i, X(t)) = \{j \mid |x_i(t) - x_j(t)| \leq \epsilon\}. \quad (9.1)$$

7. The agents update their opinions. The next period’s opinion of agent i is the average opinion of all those, which i takes seriously:

$$x_i(t+1) = \frac{1}{\#(I(i, X(t)))} \sum_{j \in I(i, X(t))} x_j(t). \quad (9.2)$$

The description 1. to 7. gives some very general interpretation of the symbols $i, j, n, t, I, x_i(t), X(t)$ which otherwise would be terms in a pure and ‘naked’ formalism. We get specific interpretations if we think of certain contexts in which the mechanisms described by equations (9.1) and (9.2) might be at work. Some such (partially overlapping) interpretations are the following:

- *Expert interpretation*: There is a group of experts on something. Each expert has an opinion on the topic under discussion, for instance the probability of a certain type of accident. Nobody is totally sure that he/she is totally right. Within bounds everybody is willing to adjust and to adapt: An expert i considers as competent all experts j that are not too far away, i.e. for which $|x_i(t) - x_j(t)| \leq \epsilon$. The updated compromise opinion is the arithmetic mean over all opinions within i 's confidence interval. Such opinion revisions produce a new opinion distribution which may lead to further revisions of opinions, and so on.¹
- *Compromise interpretation*: There is a group of people that exchange views which can reasonably be described by real-valued numbers.² For reasons as uncertainty, respect for others that seem to be as reasonable as oneself, an interest in a compromise, a preference for conformity, or due to some social pressure, everybody is, at least in principle, willing to compromise with others.³ However, the willingness to compromise is bounded: An agent i with view $x_i(t)$ is willing to compromise with agents j with view $x_j(t)$ iff j 's view is not too far away, i.e. for which $|x_i(t) - x_j(t)| \leq \epsilon$. Then averaging.
- *Social media interpretation*: There is a digital platform with a central algorithmic coordination that brings together with user i those users j whose opinions x_j are not too far away from i 's opinion x_i . Then averaging. In such a context ϵ is the 'distance tolerance' of a *centrally* organised *filter bubble* (Pariser 2011). Another variant is a digital platform that allows all users j to send their opinion x_j to all other users i . As a receiver, user i reads only opinions that are not too far away from i 's opinion x_i . Then averaging. Under this interpretation, ϵ is the distance tolerance of a *decentralised echo chamber* (Sunstein 2017).

This is not a complete list of possible interpretations. We could also use the BC mechanism as a kind of *aggregation device*, and look for the smallest ϵ that, for a given distribution of start values, leads to a consensus.⁴ From a *normative* point of view, we could use the model to analyze the effects of different recommendations

¹ Obviously, that is the procedural format of so-called *Delphi studies*.

² That includes much more than opinions on the probability of any quantitative or qualitative proposition. As long as one can reasonably normalise the range of possible opinions to the unit interval, opinions could regard any real-valued quantitative problem. The opinions could express the intensity or importance of a wish (though only under the condition of intersubjective comparability). The exchange of opinions might be about moral praiseworthiness (0: extremely bad; 0.5: neutral; 1: extremely good). Or the opinions could regard a budget share. *Not* covered are non-continuous opinions (for instance, discrete or binary).

³ The overview article Flache et al. (2017) mentions and shortly describes many empirical findings from different fields that I interpret as supporting evidence.

⁴ This interpretation gives the BC model a similar status as it was originally claimed for the Lehrer/Wagner model in Lehrer and Wagner (1981).

for how to resolve peer disagreement and other epistemic problems (Douven 2010; Douven and Kelp 2011; Douven and Wenmakers 2017). A *technical*, namely control perspective is taken in Hegselmann et al. (2015).⁵

9.1.2 Basic features of the BC dynamics

For certain ϵ -values the BC dynamics is trivial: If $\epsilon = 0$, then all agents stick to their opinions forever; if $\epsilon = 1$ then all agents' opinion is the arithmetic mean of the start profile $X(0)$ from $t = 1$ onwards. But what's about the region $0 < \epsilon < 1$? Figure 9.1 shows the dynamics for $\epsilon = 0.05, 0.20$, and 0.25 . The x -axis is the (discrete) time t ; the y -axis represents the opinion space $[0, 1]$. The grey graphs show the trajectories of the opinions of 50 agents.

For the analysis of the dynamics two concepts will be very fruitful: First, we call an opinion profile $X(t) = x_1(t), x_2(t), \dots, x_i(t), \dots, x_n(t)$ an *ordered profile* iff $0 \leq x_1(t) \leq x_2(t) \leq \dots \leq x_i(t) \leq \dots \leq x_n(t)$. Second, we call an ordered profile an ϵ -*profile* iff for all $i = 2, \dots, n$ it holds that $x_{i+1}(t) - x_i(t) \leq \epsilon$. Thus, in an ϵ -profile neighbouring opinions $x_i(t), x_{i+1}(t)$ mutually influence each other. In Figure 9.1 vertical grey lines between *two neighbouring opinions* indicate that the distance between the two is $\leq \epsilon$. As a consequence, a continuous vertical line from $x_1(t)$ to $x_{50}(t)$ means that $X(t)$ is an ϵ -profile in the sense defined above.

For an explanation of important effects, we start with the dynamics in the middle of Figure 9.1. That is the case $\epsilon = 0.2$. Careful inspection, time step by time step, shows several effects:

- Extreme opinions are under a one-sided influence and move direction center. Therefore, the range of the profile starts to *shrink*.
- At the extremes of the shrinking profile, the opinions *condense*.
- Condensed regions *attract* opinions from less populated areas within their ϵ -reach: In the center opinions > 0.5 start to move upwards, opinions < 0.5 start to move downwards.
- The ϵ -profile *splits* in $t = 5$. From now on the split sub-profiles constitute different 'opinion worlds', i.e. two communities without any influence on each other.

⁵ In substance, the model was described for the first time in Krause (1997). In a conference presentation in 1998, the naming "bounded confidence" was used for the first time; see Krause (2000). A first comprehensive analytical and computational analysis of the BC model was given in Hegselmann and Krause (2002). For the history of the BC model and a systematic classification of alternative models of opinion dynamics see Sections 1–3 of Hegselmann and Krause (2002). The closest relative of the BC model is the model in Deffuant et al. (2000); for the similarity see Urbig et al. (2008). For overviews on other, later, and future developments see Lorenz (2007), Xia et al. (2011), Sirbu et al. (2017), Flache et al. (2017).

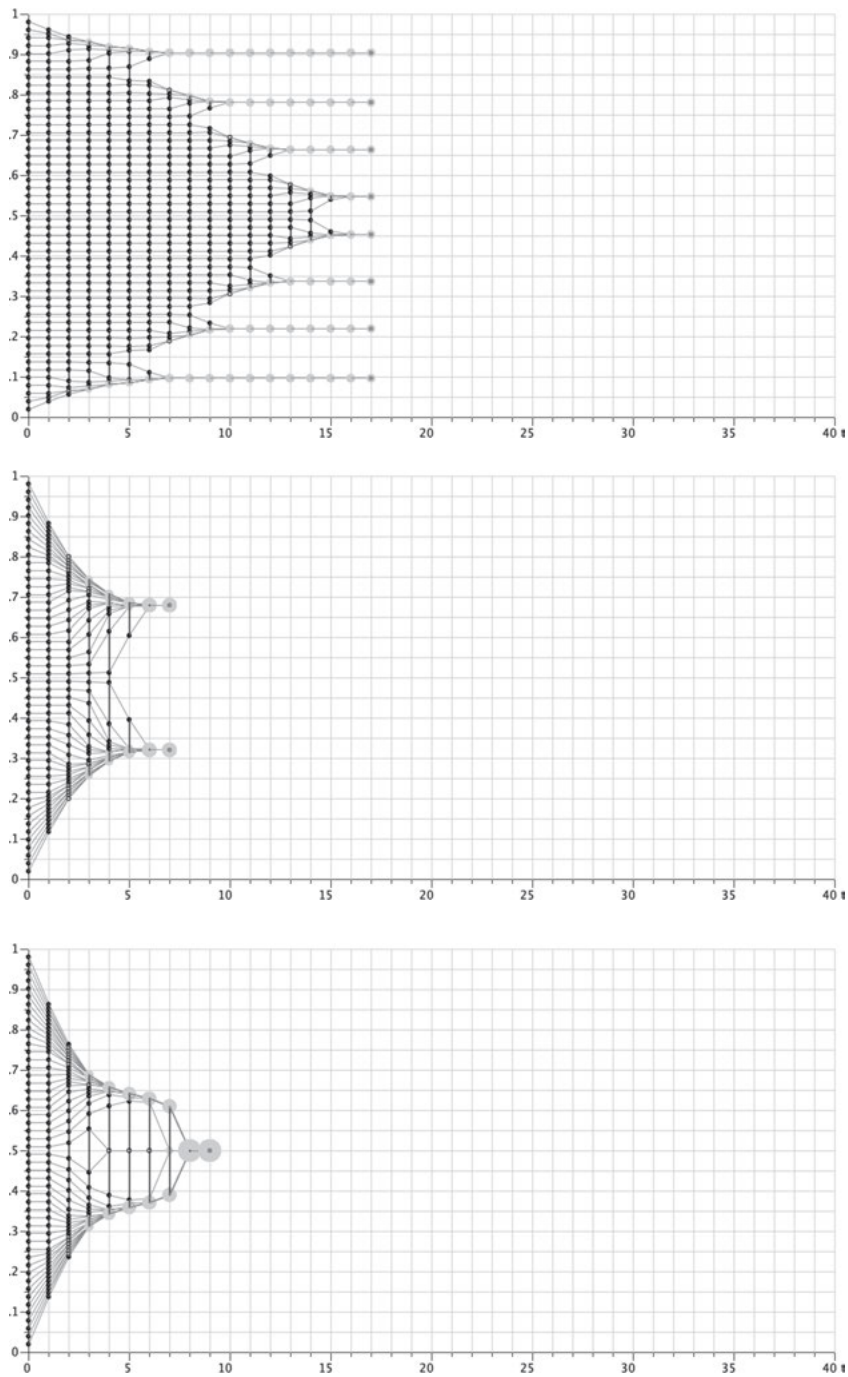


Figure 9.1: BC dynamics for 50 agents.
Notes: Top: $\epsilon = 0.05$. Middle: $\epsilon = 0.2$. Bottom: $\epsilon = 0.25$. Filled black or grey circles indicate the size of a cluster. A black circle is just one agent. A black circle with an inner white circle represents two agents. Grey circles are scaled according to their cluster size.

- In the two split off sub-profiles opinions *contract*. In $t = 5$, in the two sub-profiles all opinions have all opinions within their confidence interval. Therefore, in $t = 6$ in each sub-profile all opinions merge into one and the same opinion.
- As a consequence, we get $X(7) = X(6)$. The dynamics is *stabilized* in the sense that $X(t) = X(t + 1)$.

The dynamics in Figure 9.1, middle, ends up with two camps, one to the left, one to the right of the center of the opinion space – a kind of *polarization*. Figure 9.1, top, shows a dynamics based upon the much smaller $\epsilon = 0.05$. The ϵ -profile splits twice in period 6, leaving behind a big ϵ -sub-profile in the center that splits again in period 9. Some more splits occur some periods later. Finally 8 clusters of opinions survive. Let's call such a pattern *plurality*. Figure 9.1, bottom, is based upon a confidence interval $\epsilon = 0.25$. In this case, the ϵ -profile never splits. A cluster of two opinions in the center contracts outer opinions such that, finally, all opinions merge into one – *consensus*. In all cases, the final pattern of fragmentation – plurality, polarization, or consensus – is brought about by shrinking, condensing, attracting, contracting, splits, and mergers. And it is always brought about in *finite* time. For the latter result, simulations are not necessary: for a rigorous proof see Hegselmann and Krause (2002, Theorem 6).

The differences in the three examples in Figure 9.1 are all due to different ϵ -values. The start profile $X(0)$ is *always the same*. It is a profile for which it holds that

$$x_i(0) = \frac{i}{n+1}, \forall i = 1, \dots, n. \quad (9.3)$$

Such a profile is at the same time very specific *and* representative: The i th opinion is exactly there, where it will be at the average over infinitely repeated draws of n opinions that are *uniformly* distributed on the unit interval. Thus, equation (9.3) gives the *expected value* of the i th opinion of a uniform distribution of n opinions. We refer to that type of start distribution as the *expected value distribution*. It is a kind of *deterministic idealization* of a certain random distribution.⁶ The expected value distribution generates a start profile in which the distances between two neighbouring opinions are always the same: the start profile is *equidistant*.

In all what follows, we will rely on *one and the same* expected value start distribution with $n = 50$. For that constant start distribution we will analyze the effects of, for instance, stepwise increasing ϵ -values. That will be done by careful inspections and comparisons of *single runs*. On purpose, *programmatically*, we deviate from the

⁶ We can do the same with regard to other types of random distributions. However, one has to derive the equations that then correspond to equation (3). The equidistance is due to the uniform distribution that is deterministically idealized by equation (3). If, for example, we do the same with a normal distribution, the corresponding expected value distribution would *not* be equidistant. The relevant discipline is *order statistics*.

usual practice to run, firstly, a major number of random initialisations, and then, secondly, to do some statistics on the runs. Our approach will blind us to effects that depend largely on the randomness of initialisations.⁷ However, our approach may expose directly effects that are otherwise hidden in averages. Anyhow, I postpone the discussion of advantages and disadvantages to the concluding Section 9.4.

9.1.3 A warning: The model's simplicity is deceptive

The three dynamics in Figure 9.1 suggest a more general idea about the confidence level ϵ : The number of finally surviving and stabilized clusters decreases as ϵ increases. A first diagram – we will call it an ϵ -*diagram* (see Figure 9.2) – supports that idea.

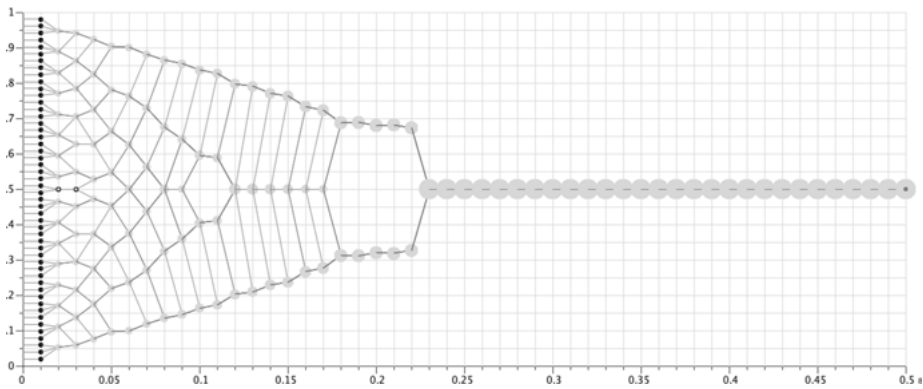


Figure 9.2: ϵ -diagram for an expected value start distribution with $n = 50$.

Notes: On the x -axes ϵ increases with step size 0.01. For a given ϵ , the y -axis indicates the stabilized end positions of opinions. The lines connect ranks in the ordered profiles. Different sizes of grey circles indicate the size of an opinion cluster. A small black circle is a cluster with just one opinion. A small black circle with a white inner circle indicates a cluster of two opinions.

An ϵ -diagram visualizes for *one and the same* start distribution $X(0)$ the effects of a stepwise increasing ϵ on the final, completely stabilized cluster structure. Figure 9.2 is an ϵ -diagram. The x -axis does *not* represent time rather than increasing values of ϵ . For each ϵ value, we run the dynamics until it is stabilized, i.e. $X(t+1) = X(t)$.

⁷ For instance, whatever n , our expected value start distribution generates *equidistant* start profiles – random start distributions do *not*. But that may have consequences for the number of finally surviving clusters.

The y-axes of an ϵ -diagram is used to display the stabilized positions. Lines connect the positions of ranks in the ordered profiles.

The lessons from the ϵ -diagram in Figure 9.2 seems to be very clear: For confidence intervals in the range $[0, 0.17]$ the number of final stabilized clusters monotonically decreases from 50 to 3 – *plurality*; in the range $[0.18, 0.22]$ we get *polarization*; for $\epsilon \geq 0.23$ the final result is always *consensus*. Thus, the general idea about the effects of the confidence level *seems* to be: Given the same, constant start distribution, for increasing values of the confidence level, the number of finally stabilized clusters decreases monotonically; we have a monotonic transition from plurality to polarization, and from there to consensus.

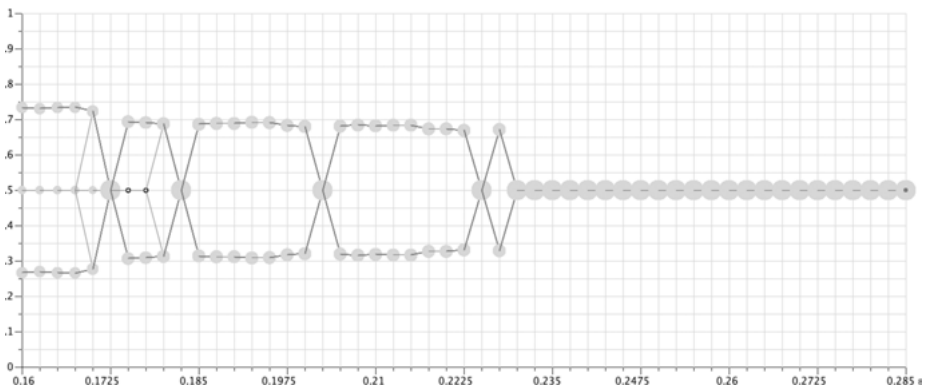


Figure 9.3: ϵ -diagram for an expected value start distribution with $n = 50$.

Notes: Different from Figure 9.2, ϵ increases on the x-axes with step size 0.0025; start is at $\epsilon = 0.16$. For other details see the caption of Figure 9.2.

Intuitive as it may be, the general idea is actually wrong, and Figure 9.3 makes that very clear.⁸ As Figure 9.2, Figure 9.3 is an ϵ -diagram for an expected value start distribution with $n = 50$. The only difference is the step size of ϵ : it is now much smaller, namely 0.0025. The smaller step size reveals that the transition from polarization to consensus is not monotonic: A consensus that is reached for a certain ϵ , may fall apart under a greater ϵ . There is always an ϵ^* such that for all $\epsilon \geq \epsilon^*$, consensus is guaranteed.⁹ But an actual ϵ that leads to consensus does not necessarily have to be ϵ^* . The non-monotonicity is not confined to the transition from two to one final clusters. And it is not an effect (or artifact) of the equidistant expected

⁸ Wedin and Hegarty (2015: 2016) write about the model: “The update rule is certainly simple to formulate, though the simplicity is deceptive.”

⁹ This is an *observation*, not a proof. For $\epsilon = 1$ we get trivially a consensus. But, as a matter of fact, normally ϵ^* is much smaller than 1.

value start distribution: the same type of non-monotonicity holds for random start distributions as well.¹⁰

9.2 Extension one: Polarization by biased confidence

The confidence intervals that we considered so far were symmetric: whatever the total size of the interval, the left and the right part were always of equal size. But what if individuals that hold a more left [more right] opinion ‘listen’ more into the left [right] direction – and the more to the left [right] their opinions are, the more so. The idea suggests an opinion dependent *asymmetry* of the following type: The more left [right] an opinion is, the more the confidence interval is *biased* direction left [right]. For opinions closer to the center of the opinion space, the bias is less pronounced. Only for opinions exactly in the center the confidence interval is symmetric.

9.2.1 Modelling biased confidence

To model the intuitive idea precisely, we divide a confidence interval of any size into a left and right proportion β_l and β_r , such that $0 \leq \beta_l, \beta_r \leq 1$ and $\beta_l + \beta_r = 1$. Following our intuition stated above, the values of β_l should be given by a monotonically decreasing function $f(x)$, defined over our opinion space $[0, 1]$. That, then, allows us to get β_r as $1 - \beta_l$. Since we request symmetry of the confidence interval for $x = 0.5$, it should hold that $\beta_l(0.5) = 0.5$.

One can get all what we want by the simple linear function

$$\beta_l(x) = mx + \frac{1-m}{2}, \quad (9.4)$$

where m , the slope of the function, controls the strength of the bias. A reasonable range for m is $[-1, 0]$. Figure 9.4 shows the resulting functions. For all values of m , the function rotates around $\langle 0.5, 0.5 \rangle$. The meaningful functions are between the dashed graphs with the slope $m = -1$ and $m = 0$. For the strongest bias $m = -1$, the most extreme leftist [rightist] opinion, that is $x = 0$ [$x = 1$], the right [left] portion of the confidence interval is 0. For $m = 0$ the opinion dependent bias disappears: For all opinions x it holds that $\beta_l = \beta_r = 0.5$ – the confidence interval is symmetric. The black graph is an example for a bias in between the extreme values for m . Given the black

¹⁰ Note that for any type of constant start profile (e.g. random or expected value), the width of an opinion profile does also not decrease monotonically with an increasing confidence level.

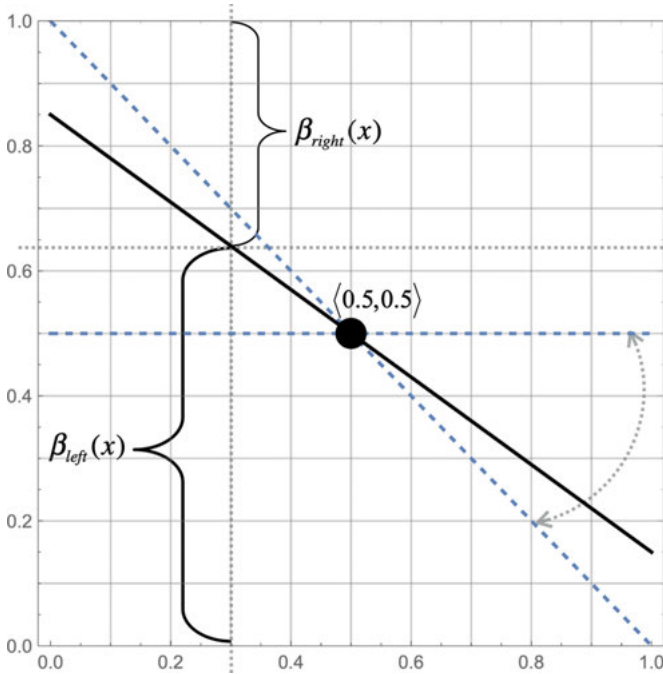


Figure 9.4: Opinion dependent bias.

Notes: The x -axis represents the opinions. The slope of the black graph gives the strength of the opinion dependent bias m (with $-1 \leq m \leq 0$) in equation (9.4). For an opinion x , $f(x)$ on the y -axis is the left proportion of the total size of the given confidence interval. $1 - f(x)$ is the right proportion.

graph with $m = -0.7$, we get for an opinion $x = 0.3$ the proportions $\beta_l(0.3) = 0.64$ and, correspondingly, $\beta_r(0.3) = 0.36$. Note: whatever the value of m , for the center opinion $x = 0.5$ we always get $\beta_l(0.5) = \beta_r(0.5) = 0.5$.¹¹

Figure 9.5, top, shows a polarization – two opinion camps, one to the left, one to the right of the center of the opinion space. It is a polarization based upon a *symmetric* confidence interval. As already noticed in Section 9.1.2, for certain regions of ϵ , polarization may evolve in the simple BC model. However, that is a rather *weak* polarization: the final distance between the two camps is only ≈ 0.36 . The two camps are much closer to the center opinion 0.5 than they are to the extremes 0 and 1, respectively. A severe opinion dependent bias, for instance $m = -1$, produces polarization of another quality. Figure 9.5 shows the difference: The dynamics in Figure 9.5,

¹¹ In Hegselmann and Krause (2002) we used the function $\beta_l(x) = mx + \frac{1-m}{2}$, defined $\beta_r(x)$ by $1 - \beta_l(x)$, and used the range $m \in [0, 1]$. As a consequence, the most severe bias is $+1$, while above it is -1 . The only reason for the change is ‘psychology’: I find it more intuitive that a function over opinions ‘from the left to the right’ gives the *left* portion of the confidence interval. Anyhow, the old and the new approach are equivalent.

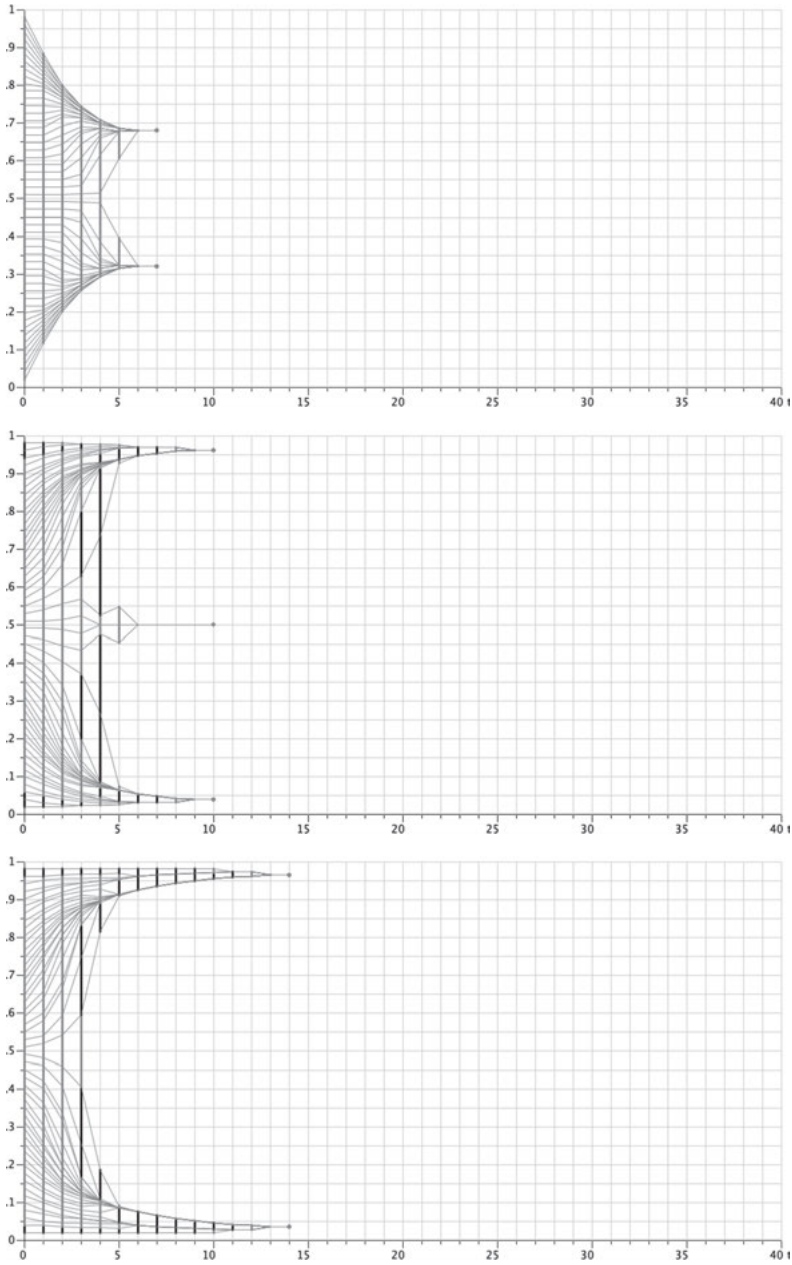


Figure 9.5: Effects of an opinion dependent bias.

Notes: *Top:* No opinion dependent bias, $m = 0$; $\epsilon_l = \epsilon_r = 0.2$ (symmetry). *Middle:* Opinion dependent bias $m = -1$, $\epsilon_{total} = 0.4$. *Bottom:* Bias $m = -1$, $\epsilon_{total} = 0.5$. Vertical *grey* lines indicate mutual influence between $x_i(t)$ and $x_{i+1}(t)$ in the profile (ascending order); black vertical lines indicate a *one-sided ϵ -split* in the opinion profile: only the more extreme of the two neighboring opinions influences the less extreme.

middle, and Figure 9.5, bottom, are both driven by the maximum bias $m = -1$. In both figures, the two opposite camps end up almost at the lower and upper bound of the opinion space: their distance is ≈ 0.92 . Thus, a severe opinion dependent bias, produced a *strong* polarization: two camps at the opposite ends of the opinion space – and the center almost or entirely empty.

How comes? Two effects contribute. *First*, under symmetric confidence, those at the extremes are under a one-sided influence of less extreme agents. Their influence drives more extreme opinions direction center. With an increasing strength of the bias, the drive direction center *disappears*, or, depending upon the total size of the confidence interval, is significantly *weakened* at the extremes. (Whatever that size, under the maximum bias $m = -1$, the most extreme opinions, do not move at all.) The *second* effect works in the opposite direction. Under symmetric confidence it always holds: If agent i influences j , then j influences i as well. In contrast, under asymmetric confidence influence may be *one-sided*. And if so, then it is a very specific one-sidedness: only the less extreme opinion is influenced by the more extreme opinion. Together the two effects generate a drift to the extremes.

In Figure 9.5 vertical lines indicate the type of influence between neighbouring opinions.¹² Grey vertical lines between neighboring opinions indicate mutual influence between $x_i(t)$ and $x_{i+1}(t)$. Black vertical lines indicate one-sided influence; *either* only $x_i(t)$ influences $x_{i+1}(t)$ *or* only $x_{i+1}(t)$ influences $x_i(t)$. Wherever a vertical line is black, we have a *one-sided ϵ -split* in the opinion profile: only the more extreme of the two neighbouring opinions influences the less extreme – no moderating influence in the other direction. Thus, above 0.5, black vertical lines connect opinions for which only $x_{i+1}(t)$ influences $x_i(t)$; below 0.5, only $x_{i-1}(t)$ influences $x_i(t)$. Figure 9.5 demonstrates a remarkable point: Different from the two-sided ϵ -splits that are definitive splits for ever, one-sided splits may after some periods close again. In the meantime they generate an unmoderated pull to the more extreme position.

9.2.2 Strange effects

At this point, a generalising intuition suggests itself: As an opinion dependent bias m stepwise increases, a weak polarization under symmetric confidence will become more and more severe – monotonically increasing the distance between the two camps, that, at opposite ends, get closer and closer to the bounds of the opinion

¹² Recall, the profiles are always ordered in an ascending order.

space. And indeed, for m -values from certain *partial sections* of $[-1, 0]$ that is true. But in general, the generalisation is definitively wrong.¹³

The refutation of the generalising intuition is given in Figure 9.6. As in the ϵ -diagrams in Figures 9.2 and 9.3, the y -axis shows stabilized end positions of opinions. Different from Figures 9.2 and 9.3, the y -axis represents the bias m . In steps of -0.01 , the bias gets stronger and stronger. It starts with $m=0$ (no bias at all) and ends with $m=-0.5$. Everything else (expected value start distribution with 50 agents, a confidence interval of a total size 0.4) is kept constant.¹⁴

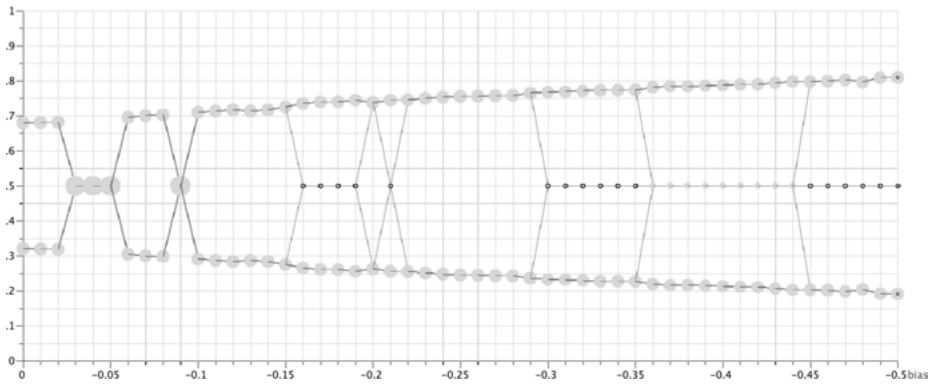


Figure 9.6: Stepwise stronger bias for the same start distribution.

Notes: On the x -axis the bias m gets stronger in steps of size -0.01 . For a given bias, the y -axis indicates the stabilized end positions of opinions. The lines connect ranks in the ordered profiles. Different sizes of grey circles indicate the size of an opinion cluster. A small black circle with a white inner circle is a cluster of two opinions. All runs start with the same expected value start distribution ($n=50$). The total size of the confidence interval is always 0.4.

The most perplexing result is that an opinion dependent bias may lead to *consensus* when, for the same start distribution, a symmetric confidence interval of the same total size leads to polarization. For an expected value start distribution with $n=50$ that is the case for $m=-0.03, -0.04, -0.05, -0.09$. Since all the underlying runs of Figure 9.6 are uniquely initialized, we can directly look into the single runs that produce the perplexing effect. Figure 9.7 does that for $m=-0.03$.

¹³ The relevant Section 4.2.2 of Hegselmann and Krause (2002) did (luckily enough) *not* claim that the generalisation is true, but – based upon that article and the underlying research – I believed the generalisation to be true until very recently.

¹⁴ We might refer to this type of diagrams as *ceteris-paribus*-diagrams (CP-diagrams). They show the final stable results of runs that are uniquely initialized; then everything is kept constant – except for *one* parameter.

The cause of the unexpected effect is visible in Figure 9.7: it is the evolution of a small cluster of just two opinions in the center of the opinion space. That cluster,

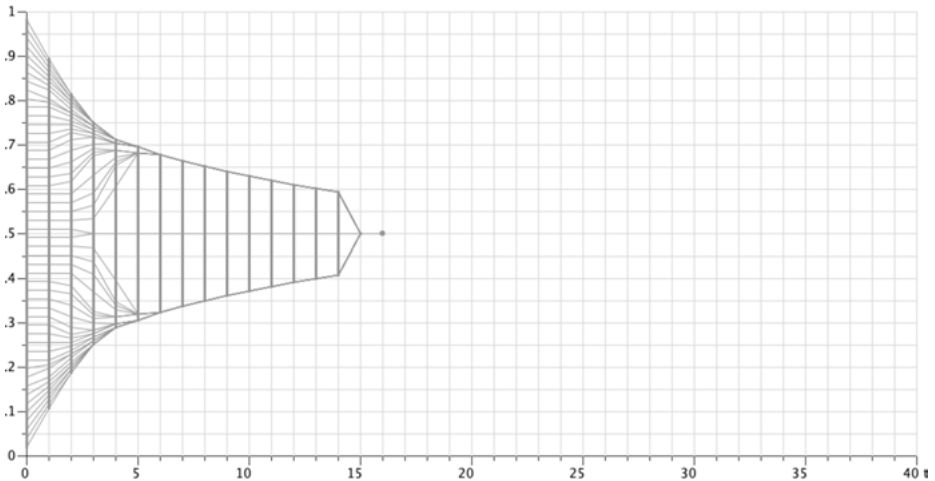


Figure 9.7: Consensus by biased confidence.

Notes: For an opinion dependent bias $m = -0.03$ the dynamics leads to consensus, while (for the same start profile and the same total size of ϵ) symmetry of the confidence interval leads to polarization (see Figure 9.5, top).

then, is a *bridge* between the two outer opinion camps and pulls them direction center. Nowhere in Figure 9.7 we see, indicated by a *black* vertical line, a one-sided ϵ -split – the perplexing results occurs without such splits. The same is true for all other cases of consensus in Figure 9.6.¹⁵

Careful inspection of Figure 9.6 reveals a further type of non-monotonicity: It is only a tiny margin, but the range of the final profile for $m = -0.20$ is *smaller* than for $m = -0.19$; in a certain sense, locally, a stronger opinion dependent bias has lead to a less extreme polarization. The same effect occurs for several other values of the bias m . In Figure 9.6 the bias gets stronger by a step size of -0.01 . For both types of non-monotonicity (as to the number of finally stable clusters, and the final width of the profile), a smaller step size would unveil even more complicated non-monotonic structures. Obviously, an understanding of the polarization dynamics includes the task to understand surprising sensitivities and non-monotonicities.

¹⁵ In the range $m = -0.5, \dots, -1$, I never found a case of consensus.

9.3 Extension two: A group of radicals enters the field

Let's now assume that there is a group of radicals. They all have the same opinion R , more or less close to the upper bound of the unit interval, for example $R = 0.9$, or even holding the most extreme position $R = 1.0$. (Alternatively, we might locate the radical position R close to the lower bound of our opinion space. Whatever we do, for the following it does not matter.) The radicals stick to their opinion – and that for ever¹⁶:

$$R(t+1) = R(t) = R. \quad (9.5)$$

The size of the radical group may matter. We refer to the number of radicals by $\#_R$. Additionally we have the *normals*, i.e. agents as we know them from the simple BC model. They have opinions from the interval $[0, 1]$, they all have a strictly positive, constant, and symmetric $\epsilon > 0$. Normals update according to equation (9.2). Now the modification comes: Whenever the radicals are in a normal agent i 's confidence interval, that is whenever $|x_i(t) - R| \leq \epsilon$, then *the whole group* of radicals is in $I(i, X(t))$. Since the radical group has $\#_R$ members, the radical position R is $\#_R$ -times in $I(i, X(t))$.

Figure 9.8 shows a single run for 5 radicals with the radical position $R = 0.9$; the black horizontal line is their trajectory. They interact with 50 normals (expected value start distribution). The normals' confidence interval is $\epsilon = 0.2$. Without radicals,

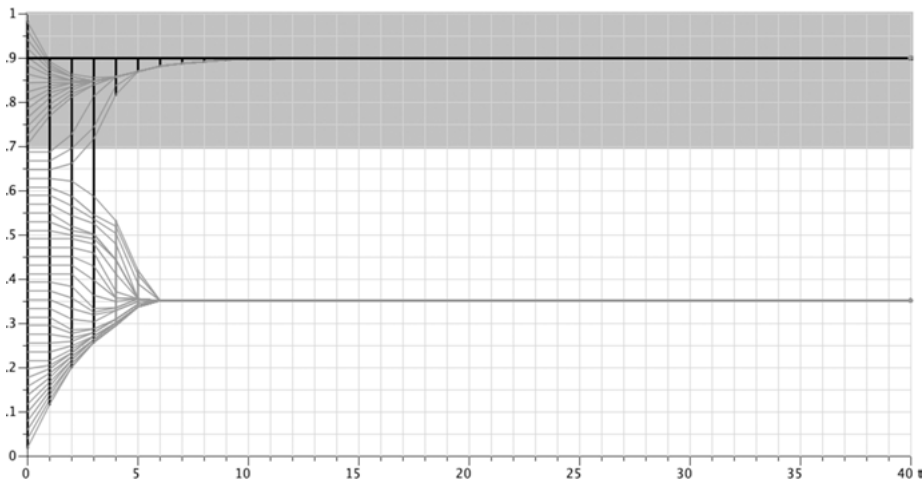


Figure 9.8: Normals and radicals.

Notes: 50 normals (expected value start distribution, $\epsilon = 0.2$) and 5 radicals ($R = 0.9$). Vertical black lines indicate a chain of direct or indirect influence of radicals on normals.

¹⁶ That is equivalent to assuming that their confidence interval is zero.

the opinion dynamics among the normals would be the dynamics in Figure 9.5, top. The dark grey area in Figure 9.8 indicates that part of the opinion space, in which all normals, given the size of their confidence interval, are under the *direct* influence of the radicals. Black vertical lines indicate the existence and length of a chain of direct or *indirect* influence of radicals on normals: Normals in the dark grey area are directly influenced by the radicals. But the radicals' influence does not end there. A normal j outside that area is indirectly influenced by a normal i inside the area of direct radical influence if $|x_i(t) - x_j(t)| \leq \epsilon$. Agent j , then, may influence other agents k outside the area of direct radical influence with opinions not further away than ϵ , and so forth. Figure 9.8 shows a far-reaching indirect influence of the radical group in the first periods: The chain of radical influence pervades the whole opinion profile, i.e. the radicals influence *all* normals. In period 4 that chain breaks. An upper part of the opinion profile converges towards the radical position. Below, the normals end up (obviously in finite time!) in a cluster. That cluster is far away from the radical position R . However, compared to the dynamics without the 5 radicals (see Figure 9.5, top), the lower cluster's final position is a bit shifted in the direction of the radical position. Obviously indirect radical influence matters.

It is very natural to think, that final numbers of radicalized normals crucially depend upon the number of radicals compared to the number of normals (hopefully only the ratio matters), the confidence level ϵ , and the radicals' position R . Under this working hypothesis, our model has only few parameters and we should be able to answer our questions for major parts of the parameter space. One possible simulation strategy, then, is the following: Let's assume we have 50 normal agents and the most extreme radical position that is possible, i.e. $R = 1.0$. Two parameters are left: The number of radicals, and the confidence level ϵ . Now we put a grid on the two dimensional parameter space: In 50 steps of size 0.01 the confidence level of normals increases from 0.01 to 0.5 on the x -axis. On the y -axis the number of radicals increases in 50 steps from 1 to 50 (then the group of radicals has as many members as the group of normals). We compute the runs for each of the 50×50 parameter constellations $\langle \epsilon, \#_R \rangle$. A run is considered stabilized, iff the opinion profiles $X(t)$ and $X(t+1)$ are almost the same. More precisely, we stop a run if for *all* agents i it holds that $|x_i(t+1) - x_i(t)| \leq 10^{-5}$. As to statistics, we will focus on *one* number only: the number of normals that finally hold an almost radical position. And we consider a normal agent i 's opinion as almost radical iff $|x_i(t) - R| \leq 10^{-3}$.

Again, for all runs in the whole parameter space we will use one and the same expected value start distribution as given by equation (9.3). As a consequence, there isn't any randomness in our analysis. If we find in the grid of 50×50 parameter constellations $\langle \epsilon, \#_R \rangle$ some interesting effects, then we can go straight into the *unique single runs* that generated perplexing macro-effects.

There are other models of radicalization, for instance Deffuant et al. (2002) or Baumann, Betz, and Cramm (2014). Often, they are more complicated than ours, or do not allow for the type of analysis that we want to apply. We go for extreme

simplicity and aim at a complete understanding, including the micro level that brings about the patterns on the macro level, for instance the numbers of finally radicalized agents.¹⁷

9.3.1 A bunch of surprising effects

Result of the computation of single runs for the 50×50 parameter constellations $\langle \epsilon, \#_R \rangle$ is an array of numbers. It is much easier to detect patterns and structures in *colored landscapes* rather than in an array of numbers (even if they are only integers – as in our case). Therefore (as a kind of phase diagram) Figure 9.9 shows, indicated by color, the number of normals that finally end up at the radical position $R=1.0$. For ease of reference, we refer by the capital letters A, B, C, D, E, and F to certain regions of the parameter space, as they are partitioned by the black lines (two vertical, one horizontal).

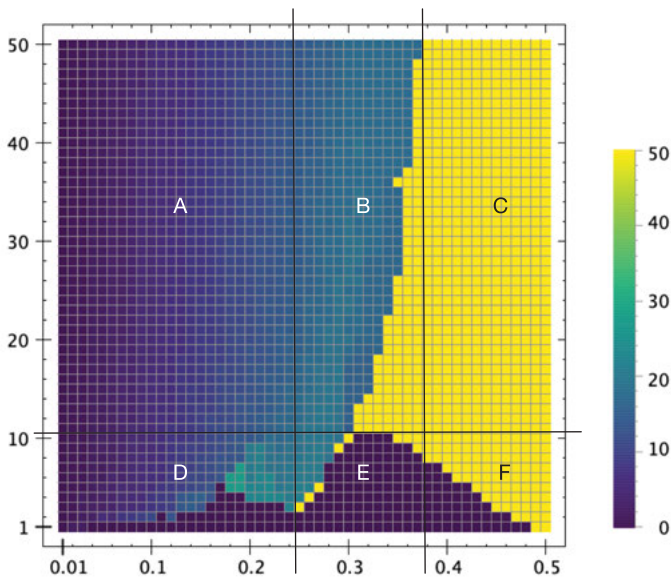


Figure 9.9: Radicalization landscape for an expected value start distribution.

Notes: x-axis: the confidence level increases in 50 steps of size 0.01 from 0.01 to 0.5. y-axis: the number of radicals increases from 1 to 50. Colors indicate the number of normals that end up at the radical position which is here assumed to be $R=1.0$. The total number of normals is always 50.

¹⁷ The analysis presented here is extended in Douven and Hegselmann (2020).

On the y -axes the number of radicals increases stepwise. Therefore, sudden dramatic color changes in vertical direction are dramatic changes in the number of radicalized normals that – *ceteris paribus* – are caused by just *one* more radical. Correspondingly, a dramatic color change in horizontal direction is – *ceteris paribus* – a dramatic change in the number of radicalized normals caused by a tiny increase of ϵ by $1/100$.

In the following we inspect region-wise our parameter space. We start with the three vertical regions. Our question is, how the number of normals that end up at the radical position, depends upon the number of radicals. Then an inspection of the two horizontal region follows. Our question there, is, how the radicalization of normals depends upon confidence levels.

1. In the region $F \cup C$, i.e. a region with *higher* confidence levels ϵ , the number of radicalized normals *monotonically* increases as $\#_R$ increases. But for all $\epsilon < 0.49$, there is a sudden jump: One more radical, and the number of radicalized normals jumps from none to all. Obviously there is an ϵ depending threshold $\#_R^*$ of radicals, such that, first, for that threshold *no* normal ends up at the radical position, while, second, for $\#_R^* + 1$ *all* normals end up at the radical position.
2. In the region $E \cup B$, i.e. a region with *middle-sized* confidence levels we find jumps of all sorts and in various directions: In region E there are – again in vertical direction – jumps from none to all: One more radical, and, instead of none, all normals end up at the radical position R . But, additionally, in region E and B there are jumps in the opposite direction: One *more* radical, and instead of all, *significantly less* (about half of the normals, or even less) end up radical.

By careful inspection of region $E \cup B$ in Figure 9.9 one can verify: With the exception of one of the ϵ values (the exception will be discussed later), it holds for the *middle-sized* confidence levels in region $E \cup B$:

1. For all of them exists an threshold $\#_R^*$ for none-to-all jumps.
2. For all of them exists another ϵ depending threshold $\#_R^{**}$ of radicals, such that, first, for that threshold *all* normals end up at the radical position, while, second, for $\#_R^{**} + 1$ *significantly less* normals become radical. Obviously, there is a second type of jumps, now working into the opposite direction.
3. In region E the second threshold $\#_R^{**}$ equals $\#_R^* + 1$. As a consequence, two steps of adding just one more radical causes the dramatic change from none to all, and then back to about half of the normals being radicalized.
4. For the thresholds $\#_R^{**}$ that are in region B , the jump from all to significantly fewer comes later – but it comes: There is always a number of radicals such that just one more reduces the number of radicalized normals from all to about less than the half.

There is one exception from observation (a) to (b): For step 28 on the y -axis ($\epsilon = 0.28$) there is a threshold $\#_R^{**}$, but no threshold $\#_R^*$. To be frank: I do not know

the reason. May be it is simply as it is – in the non-linear BC-dynamics often minor differences matter. The missing threshold may be a hint, that the exact position pattern of the thresholds $\#_R^*$ is a more complicated issue than it looks under our 50×50 grid of $(\epsilon, \#_R)$ parameter constellations. A grid that is finer with regard to ϵ could give an answer. And finally, bleak as it is: The missing threshold $\#_R^*$ may be the consequence of numerical problems.¹⁸

Leaving the one exception aside and summing up: $E \cup B$, a region of *middle-sized* confidence levels, is a region with sudden ups (from none to all) and downs (from all to significantly less) radicalized normals. Along certain lines in the parameter space the sensitivity to tiny changes is extreme. The predominant phenomenon is, that the number of radicalized normals is *not* monotonically increasing with an increasing number of radicals. Just one more radical may lead to much less radicalization.

Even in the smooth areas of region B the radicalization of normals is clearly *not* monotonically increasing with regard to $\#_R$. On the contrary: In the left part of area B the radicalization of normals is slightly *decreasing* as the number of radicals *increases*.

3. The region $D \cup A$ is the region of *smaller* confidence levels ϵ . Again, for an increasing $\#_R$, there are certain threshold values where jumps occur. But they are not jumps from none to all. Nevertheless, they are jumps from none to a significant proportion. In the right part of D the sudden increase is more drastic than in the left part. Again there is a striking effect: Above the jumps from none to a significant proportion, the number of radicalized normals clearly *decreases* as the number of radicals increases – *less* radicals would have *more* effect.

4. *Horizontally*, i.e. with regard to $\#_R$, we distinguish *two* regions. There is an upper region with a *major* number (or proportion) of radicals, the region $A \cup B \cup C$. It is a region with always more than 10 radicals, i.e. a radical group size of more than $1/5$ of the number of normals, or, respectively, more than $1/6$ of the whole population.¹⁹ Given such a major $\#_R$, if ϵ increases, there always exists a threshold ϵ^* such that for $\epsilon^* + 0.01$ the number of radicalized normals jumps from about $1/3$ to all. The upward jumps are compatible with monotonicity. However, careful color inspection of the area to the left of the thresholds ϵ^* clearly shows (especially clear for the middle sized confidence levels in region B) that, with increasing confidence levels, $\#_R$ – slightly and smoothly – first increases, but then decreases (again slightly and smoothly). In sum, for major numbers of radicals, as to the numbers of radicalized normals, there is no general monotonicity with respect to the size of their confidence interval.

¹⁸ For a description and discussion of the numerical problems see Hegselmann and Krause (2015: 483ff.).

¹⁹ If one counts the cells up to the horizontal line, the result is 10. Note: The y-axis' origin is 1 (and the x-axis origin is 0.01).

There are *two* regions in the whole parameter space, that behave very smoothly: the regions *A* and *C*. Both belong to the upper horizontal region with a *major* number of radicals that we inspect right now. In region *C*, that is for *higher* confidence levels, for any $\#R$ all normals end up radicalized. In region *A*, that is for *smaller* confidence levels, it is the normals' confidence level that matters – not $\#R$: The radicalization of normals increases as their ϵ increases. In the bottom right area of *A* the number of the radicals has a bit effect: The number of normals, that end up at *R*, slightly decreases as $\#R$ increases.

However, a warning side remark: Figure 9.9 shows the number of normals that become radical. We consider a normal agent *i* as “radical”, “radicalized”, “ending up at the radical position” etc. iff $|x_i(t) - R| \leq 10^{-3}$. Therefore, even if $\#R$ has (almost) no effect on the number of – in this sense – radicalized normals, it may nevertheless have (and often has) a major effect on the mean or median opinion of the normals' opinions, the cluster structure etc., which we do not analyze here.

In region *A* and *C* the number of radicals has very little or no effect on the radicalization of normals. In region *B* that is different, and $\#R$ seriously matters: The exact location (though not the existence) of the threshold ϵ^* depends upon the number of radicals: As their number increases, the jumps occur more to the right, i.e. they require higher confidence levels.

5. The lower horizontal region, i.e. $D \cup E \cup F$, is a region with *minor* numbers of radicals (not more than $1/6$ of the whole population of normals *plus* radicals). In terms of jumps it is the wildest region: In *E* and *F* we find (as in *B*) thresholds ϵ^* such that for $\epsilon^* + 0.01$ the number of radicalized normals jumps from less than a half or even none to all. But, additionally, there are values ϵ^{**} , such that for $\epsilon^{**} + 0.01$ the number of radicalized normals jumps from all to zero. The most striking point is, that in *E* both threshold values are horizontally next to each other, i.e. $\epsilon^{**} = \epsilon^* + 0.01$. Obviously, the radicalization of normals reacts in *E*, i.e. an area with both, a minor number of radicals *and* a middle-sized confidence level, extremely sensitive with regard to both initial conditions, the confidence level *and* the number of radicals.

However, there is again a conspicuity in area *E*, now in vertical direction: For all $\#R > 2$ *except* for $\#R = 6, 7, 8$ there exists a threshold ϵ^* (as defined above). Again we do not know the reason. The same considerations, as mentioned above in the corresponding case for $\epsilon = 0.28$, apply (see the second observation).

In area *D* we find for an increasing ϵ a lot of jumps in both directions: from some to none and from none to some. The jumps are less dramatic than in region *E*, but they are there. For a description one might introduce thresholds that correspond ϵ^* and ϵ^{**} but have reduced requirements.

To sum up, with respect to the confidence level, the whole region $D \cup E \cup F$ is a region of non-monotonicity, jumps up and jumps down.

What are the main results of our inspection? There are two ‘smooth’ areas in the parameter space, the areas *A* and *C*. But in *all* other areas we find the thresholds ϵ^* ,

ϵ^{**} , $\#_R^*$, and $\#_R^{**}$ (in D we observe corresponding thresholds with reduced requirements). Let's call a region of our partitioned $\langle \epsilon, \#_R \rangle$ -parameter space *wild* iff, first, we have in that region a non-monotonicity (both, decreasing *and* increasing) with respect to one or both parameters, and, second, the region is pervaded by sensitivities. Under that definition we can distinguish *two* wild regions: In *vertical* direction the region $E \cup B$, a region of middle-sized confidence levels; in *horizontal* direction the region $D \cup E \cup F$, a region of comparatively small numbers of radicals.

For all regions, wild or not, immediately “Why is it, that . . . ?”- questions arise. Why is it, that in region C neither the number of radicals, nor the confidence level has any effect on the number of radicalized normals? Whatever the specific parameter constellation in that region, *all* normals end up radical – but why? Why is it, that in region A , a region where the number of radicals is above $1/6$ of the whole population, radicalization of normals is *not very much* influenced by the number of radicals. Obviously, it is the confidence level of normals that matters – but why?

In the following we will make *expeditions* into the ‘wild’ regions of our parameter space and try to understand, how ‘in the deep’ some completely deterministic mechanisms create the wild ‘*radicalization landscape*’, that Figure 9.9 displays.

9.3.2 An expedition into the wild parameter region

Our use of always the same expected value start distribution has *two* consequences: First, whenever we want to understand what causes certain effects in our radicalization landscape, we can go directly into *unique single runs*. No statistical analysis of 100 or so randomly started runs is necessary. Second, since, additionally, all the single runs start with the same start distribution of normals, we can, by comparison of single runs, *on the level of single agents* directly observe the effects of changes of ϵ or $\#_R$. Especially, if we inspect single-run-sequences of *small stepwise changes*, we directly observe the working of the ‘forces in the deep’ that generate the surface of our radicalization landscape – and that should be a good starting point for an identification and understanding of the mechanisms that bring about the puzzling landscape.

Our expedition will be an expedition direction north that starts at $\epsilon = 0.31$. On that path are two puzzling points: First, a sudden change from none to all of the normals being radicalized (explanandum 1); second, with only a few more radicals, a sudden change from all to only about $1/3$ normals being radicalized (explanandum 2). For each of the parameter constellations that we pass going north on the 50×50 grid of $\langle \epsilon, \#_R \rangle$, we have unique single runs with the same start distribution. Therefore, we can generate a sequence of 50 pictures, one for each $\langle \epsilon, \#_R \rangle$ constellation that we pass. Each of the pictures displays the trajectories of all 50 agents. The time scale on the x-axis is always the same: 50 periods. Going strictly north implies, that ϵ is kept constant. Therefore, whatever is changing in the sequence of pictures, it is the consequence of *one* factor only: the number of radicals.

Figure 9.10 displays the *first* explanandum: The jump from none to all normals being radicalized when the number of radicals increases from 10 to 11. For an explanation we start in Figure 9.10 bottom, i.e. the dynamics under the influence of 10 radicals. As the dark grey vertical lines indicate, there is up to period 7 (we start with period 0) a chain of direct or indirect influence of the 10 radicals even on the most distant normals. Soon 3 opinion clusters emerge among the normals. The cluster in the middle functions as a bridge between the upper and the lower cluster. The upper cluster is in the dark grey area of the opinion space, and that is the area of direct influence of the radicals. Thus the upper cluster of normals is a bridge between the radical group and other two clusters of normals, which, period by period, move direction R . But that works only for a while: As the lower cluster of normals moves

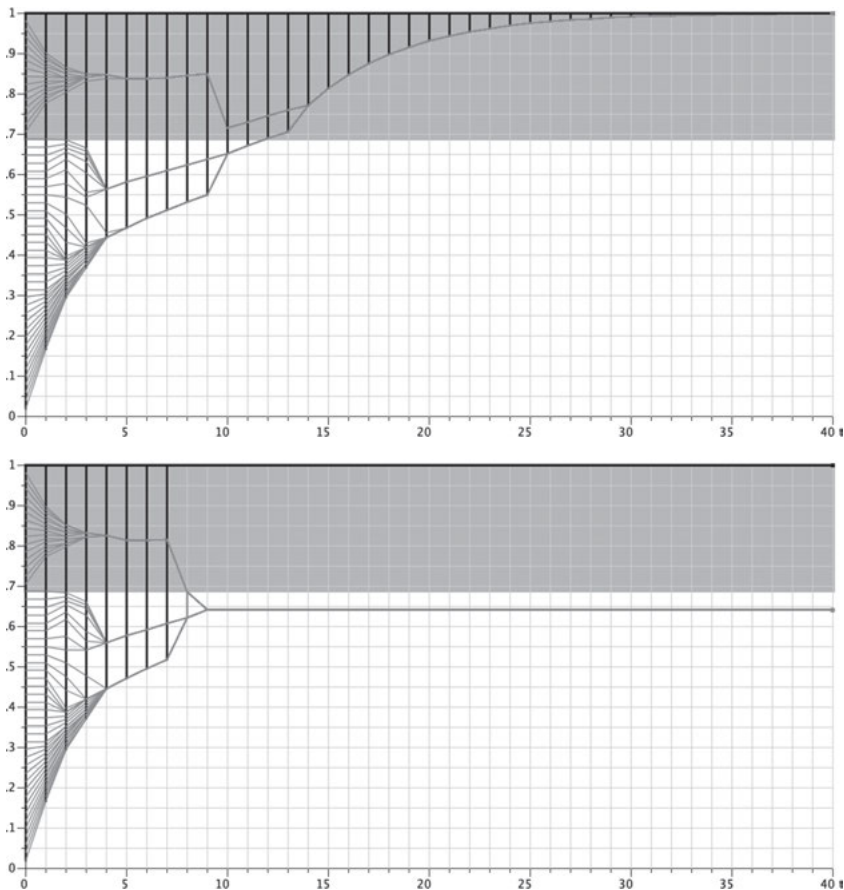


Figure 9.10: One more radical and everybody gets radicalized.

Notes: Bottom: $\#_R = 10$. Top: $\#_R = 11$. $\epsilon = 0.31$. One more radical causes a jump from none to all of the normals ending up at the radical position.

upward, at a certain point the upper normal cluster has both, the middle and the lower cluster within its confidence interval. Their combined influence on the upper cluster is strong enough to pull the upper cluster completely out of the area of direct influence of the radicals. The consequence is, that no bridge between radicals and normals exists any longer. However, the radicals had an effect on the normals: Without radicals the normals would end up at a 0.5-consensus. With 10 radicals ($1/5$ of the number of normals, $1/6$ of the whole population) it is about 0.64.

Now we add just one radical, analyze the trajectories in Figure 9.10 top, and compare it with what we see in Figure 9.10 bottom: With the one more radical, again, the three clusters of normals evolve. However, it takes a few periods more until the upper cluster gets under the direct influence of both clusters below, which, therefore, both are moving direction R a bit longer, before, then, that period comes. When it comes, the upper cluster makes – as in the case with one radical less – a steep move away from the radicals' position. But different from the case with $\#_R = 10$, the upper cluster does *not* get out of the area of direct influence of radicals. From that moment onwards everything is lost: Though now further away from the radical position, the upper cluster *continues* to function as a bridge between all other normals and the radicals. After two more periods the bridge isn't necessary any longer: Now *all* normals are under the *direct* influence of the radicals – and that is a point of no return: From now on (and as one cluster) all normals irreversibly move and converge – though in infinite time – to the radical position.

The inspection of the underlying dynamics resolves the puzzling effect in the radicalization landscape: The sudden jump from none to all is due to what we might call a *positive bridging effect*: For $\#_R = 10$ a cluster evolves that for a while functions as a bridge between the radicals and all other normals. But then the bridge breaks down and, additionally, the upper cluster gets out of the area of direct radical influence. For $\#_R = 11$ the bridge to the radicals *continues* to function until it becomes superfluous. The functioning of the bridge is critical and accounts for the difference between none or all of the normals ending up radical.

Figure 9.11 displays the *second* explanandum: The sudden *drop down* from all normals being radicalized to only $1/3$, and that by *increasing* the number of radicals from 13 to 14. We start our analysis in Figure 9.11 bottom. What we see there, is very similar to Figure 9.10 top. Based on what we saw there, we understand the positive bridging effects that are (still) at work in the dynamics in Figure 9.11, bottom. But the one more radical in Figure 9.11, top, causes dramatically different trajectories: the middle cluster of normals is somehow 'blown up': One of the former members joins the upper cluster, all others join the lower cluster. For some periods the radicals still influence even the most distant normals. But there is no evolution of a bridging cluster *in-between* the two clusters of normals. The distance between the two clusters enlarges. As a consequence, the radicals' chain of influence breaks and becomes very short afterwards. More than $2/3$ of the normals form a cluster at about 0.45, that is even slightly *lower* than the center of the opinion space.

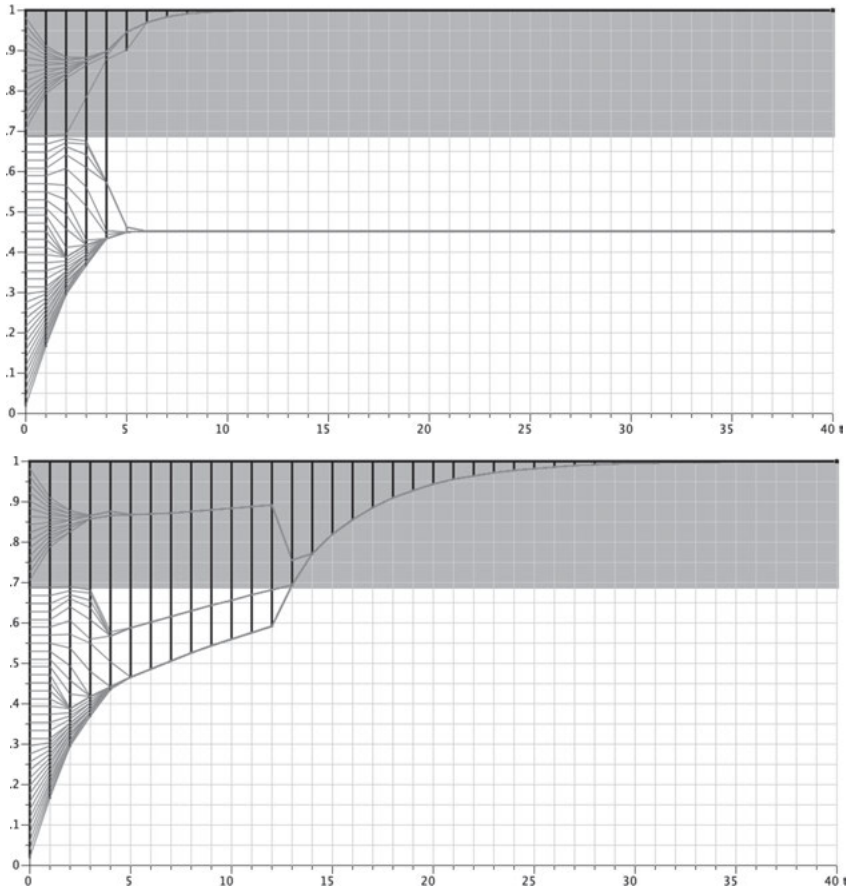


Figure 9.11: One more radical and the radicalization goes down.

Notes: Bottom: $\#_R = 13$. Top: $\#_R = 14$. $\epsilon = 0.31$. One more radical causes a jump down from all to less than one third of the normals ending up at the radical position.

Decisive for the sudden jump downwards is, that the one more radical causes a rupture in a segment of the normals' opinion profile that, without the additional radical, would have become a bridging cluster. – Obviously, there are not only positive bridging effects. In our explanation of the second explanandum a *negative* bridging effect is at work: Under one more radical a former bridge to the radicals ceases to exist. And that causes a dramatic reduction in terms of radicalized normals.

We can't analyze here all the details of the mechanics that works underneath the radicalization landscape on the route further north. Only so much: The number of radicalized normals goes up to a maximum of 19, fluctuates for a while between 18 and 19, and ends for $\#_R = 50$ with 18 radicalized normals. *All* explanations of these figures, their small range, and their fluctuations, are about the details of how the cluster, that for $\#_R = 13$ functions as a bridge between the upper and the lower

cluster of normals (see Figure 9.11, bottom), is ‘blown up’, disassembled, and ruptured into pieces, once we add one more radical, and another one, and so forth.

In a similar style we could do an expedition going east, for instance starting at $\#_R = 5$. As Figure 9.9 shows, on that path we would encounter three dramatic sudden changes that ask for explanation.²⁰ What lessons can we learn from our expeditions? The most important lessons are these:

1. A prominent role in all explanations of sudden jumps of the number of finally radicalized normals have *bridges from normals to radicals*. They require, as a kind of pier, normals (cluster or single) that, given a confidence level ϵ ,
 1. are themselves *inside* the area of direct radical influence,
 2. are *within* the confidence interval of *other* normals, that are *outside* the area of direct radical influence.

Let’s call such bridges *type-R bridges*. They are decisive for any influence of radicals outside their limited area of direct influence (which is determined by the normals’ confidence level). Type-R bridges allow for *indirect* influence of radicals on normals.

There is a second type of bridges, *bridges from normals to normals*. They require (again as a kind of pier) normals (cluster or single), that, given a confidence level ϵ ,

1. are themselves *inside* the confidence interval of at least *two other* normals (cluster or single),
2. that themselves are *outside* each other’s confidence interval.

Let’s call this type of bridge *type-N bridge*.

What we have seen, then, is, that via an uninterrupted chain of bridges, starting with a type-R bridge and then prolonged by a number of type-N bridges, the radicals may have an influence even on normals that are far away from the radical position. But, except for the group of radicals, the piers of our bridges *can move over time* – and that may *destroy* a bridge, whether of type-R or type-N. At the same time *new* piers for *new* bridges may evolve. *Lesson: Understanding the radicalization landscape is an understanding of types, evolution, and breakdown of bridges in a dynamical network.*

2. The probably most striking puzzles are sudden jumps *down* from all to none, or significantly fewer radicalized normals – and that caused by an *increasing* $\#_R$ or ϵ (as an expedition direction east would demonstrate). In explanandum 2 one more radical causes a pull upwards, which, *via* a type-R bridge, disrupts a former and essential type-N bridge. The movable pier of the type-N bridge moves steeply direction *R*. Thereby the bridging capacity, given by ϵ , is over-stretched, and the type-R-bridge breaks down (see Figure 9.11).

²⁰ Hegselmann and Krause (2015) describes that expedition in detail.

An increasing number of radicals may have the effect that the upward pull disrupts piers of type-*N* bridges and/or attracts too fast the pier of a type-*R* bridge. An increase of ϵ causes a stronger contraction. That may sweep along a former pier of a type-*R* bridge, and the former pier gets outside the area of direct radical influence. In both cases the breakdowns of bridges depends upon thresholds. Therefore they are sudden events. In both cases the breakdown of bridges may stop the radicals' influence on major fractions of normals. As to the numbers of radicalized normals, even jumps from all to none are possible. *Lesson:* We can explain the sudden jumps downward in terms of effects on type-*R* and type-*N* bridges in an opinion profile, that is exposed to two interlinked forces, that get stronger: the first pulls upwards, the second contracts the range of the profile. If the forces get stronger, they may destroy decisive bridges of influence.

3. Sudden jumps upwards are another puzzling effect. Such a jump occurs in the *explanandum* 1. With one more radical the pier of a type-*R* bridge is no longer pulled downwards outside the area of direct radical influence (see Figure 9.10). *Lesson:* Obviously we can explain the sudden jumps downwards in terms of effects of the interlinked forces for type-*R* or type-*N* bridges. If the forces get stronger, bridges that before broke down, may keep functioning, or piers for new bridges may evolve.

That are some lessons. Many questions are left open, for instance with regard to the location of the extreme sensitivities. To answer them requires many more expeditions into the wild.

9.4 Concluding remarks and future perspectives

In this article we presented two extensions of the BC Model. In Section 9.2 we introduced an opinion dependent bias. In Section 9.3 radicals entered the field. Both extensions revealed phenomena of general importance: *First*, the importance of bridges; *second*, an extreme sensitivity in certain regions of the parameter space. In these regions very small differences in the parameter values have massive effects.

With regard to *bridges*, in the polarization processes of Section 9.2, opinion clusters that had developed in the middle of the opinion space were able to pull evolving outer clusters towards the center; without the bridge in the middle, the outer clusters would belong to different opinion worlds. Conversely, the absence or disappearance of such bridges contributed decisively to polarization. In Section 9.3, bridges or chains of bridges were decisive for the influence of a group of radicals on normal BC agents. They connected the cluster of radicals with clusters of normals outside the direct influence of the radicals. The ϵ splits that we introduced in the context of the basic BC model, can also be described as the absence or disappearance of bridges. The term *bridge* in the sense as we use it here, originates from network theory. The term is explicitly introduced in the network theory classic Harary,

Norman, and Cartwright (1965). In the classical sociological application of network theory, namely in Granovetter's *The Strength of Weak Ties* (1973), the weak ties *are* bridges. The use of the term in the context of the BC model is *not* a transfer of the concept into another field: The BC dynamics can be understood as a dynamic network in which the edges connect agents (nodes) whose opinion distance is not greater than ϵ .

As to *sensitivity*, we saw in the first extension that polarization reacts extremely sensitive to small changes of a relatively weak bias. In the second extension the number of radicalized normals reacts extremely sensitive to small changes of a relatively small proportion of radicals or, respectively, small changes of a medium-sized confidence interval. The sensitivity is associated with certain non-monotonicities: a more pronounced bias does not necessarily lead to a more pronounced polarization – it can even lead to consensus; while an even stronger bias then leads to polarization again. In the second extension, more radicals or larger confidence intervals can lead to a larger, but also to a smaller number of radicalized normals.

In the first instance, this sensitivity is the sensitivity of artificial worlds created as instantiations of the BC model. If, however, with regard to relevant and critical aspects, these artificial worlds are sufficiently similar to the real world, then one would also have to reckon with all the mechanisms and effects in the real world. As a consequence, for an understanding of polarization and radicalization processes in the real world, a concentration on network bridges would be a good idea. As far as details are concerned, things could become epistemically and politically difficult in the real world: In certain parameter ranges, anyone who wants to make a prognosis on the results of an actual dynamics of opinions would have to have extremely precise knowledge of the situation, that is the actually given parameter values. The same applies to the exact explanation of a factual dynamics. A social planner who could within certain but narrow limits, manipulate parameter values, would have to know very precisely where he or she actually is in the parameter space in order to even give his or her intervention only the desired direction. In a radicalization landscape that we calculate for one and the same *random* start distribution, the difficulties would become even more severe.

All results and insights of this article are based upon one and the same expected value start distribution of 50 opinions according to equation (3). That made it easy to analyze the factors and mechanisms that cause sudden massive changes in sensitive parameter regions. But are the findings that the radicalization landscape of Figure 9.9 shows perhaps essentially due to the *equidistance* of opinions in the expected value start profile that deterministically idealizes an even random start distribution? Would this landscape have looked substantially different if one carries out 50, 100 or 1000 experiments with 50 randomly distributed opinions of normals (uniform distribution) and then averages over the number of radicalized normals? The answer is given by such random simulations carried out by Igor

Douven.²¹ As Figure 9.12 shows, the answer is: *No*, the radicalization landscape calculated in this way does not look much different. The abrupt changes are slightly blurred, but they remain abrupt – the sensitivity is preserved. In this respect, the general use of the same expected value start distribution is obviously a good representative substitute for high numbers of random experiments. The required computing time is reduced by one to three orders of magnitude. A radicalization landscape as the 50×50 parameter constellations $\langle \epsilon, \#_R \rangle$ with colors indicating the final number of radicalized normals, constitutes a major *explanandum*. If the underlying start distribution is both representative and always the same, there is a straightforward strategy to find explanations: we directly go into the single runs that produced the puzzling radicalization patterns. In our case the single run analysis made very clear that the crucial point is the emergence and the collapse of bridges (type-R or type-N bridges) between clusters.

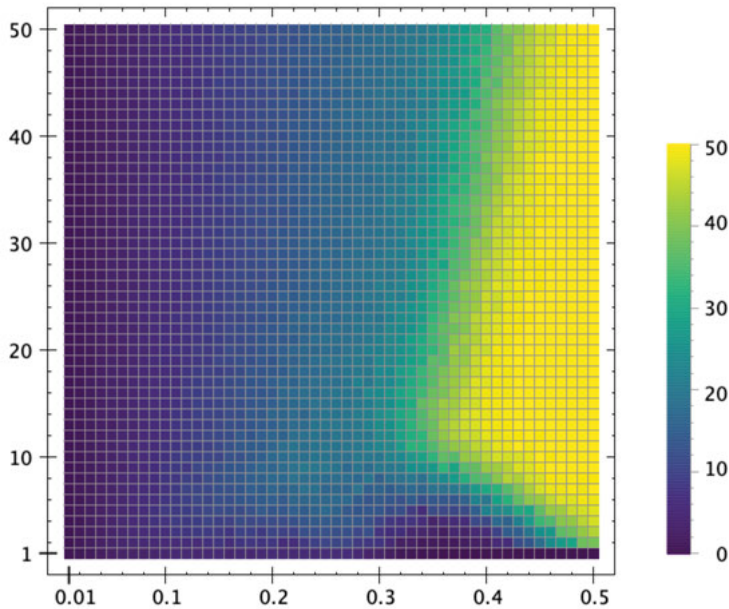


Figure 9.12: Radicalization landscape for random start distributions.

Notes: For each of the 50×50 parameter constellations $\langle \epsilon, \#_R \rangle$, Igor Douven ran 100 simulations based upon random start distributions (uniform). x-axis: the confidence level increases in 50 steps of size 0.01 from 0.01 to 0.5. y-axis: the number of radicals increases from 1 to 50. As in Figure 9.9, the radical position is $R=1$. Colors indicate the *average* number of radicalized normal.

²¹ On November 1, 2018, Igor Douven has sent to me the respective radicalization landscapes. They were computed with the newly developed, very fast *Julia* programming language.

All that does not mean that, from now on, we should forget about iterated runs based upon random start distributions. In the research strategy that I propose here, runs with random start distributions are still important, but more as a *hedging measure*: By computing as well the radicalization landscape for iterated random start distributions we hedge against the danger of artifacts, as they may be caused by expected value start distributions. The similarity between the landscape that is based upon one and the same expected-value start distribution (Figure 9.9) and the landscape that is based upon iterated runs based upon random start distributions (Figure 9.12) suggests that our single-run based explanations of puzzling phenomena regard and cover generic effects.

References

- Baurmann, Michael, Gregor Betz, and Rainer Cramm. 2014. "Meinungsdynamiken in fundamentalistischen Gruppen – Erklärungshypothesen auf der Basis von Simulationsmodellen." *Analyse und Kritik* 36 (1): 61–102.
- Deffuant, Guillaume, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. "Mixing beliefs among interacting agents." *Advances in Complex Systems* 3 (01n04): 87–98.
- Deffuant, Guillaume, Frédérik Amblard, Gerald Weisbuch, and Thierry Faure. 2002. "How can extremism prevail? A study based on the relative agreement interaction model." *Journal of Artificial Societies and Social Simulation* 5 (4).
- Douven, Igor. 2010. "Simulating peer disagreements." *Studies in History and Philosophy of Science* 41 (2): 148–157.
- Douven, Igor and Kelp, Christoph. 2011. "Truth Approximation, Social Epistemology, and Opinion Dynamics." *Erkenntnis* 75: 271–283.
- Douven, Igor and Sylvia Wenmackers. 2017. "Inference to the Best Explanation versus Bayes's Rule in a Social Setting." *British Journal for the Philosophy of Science* 68: 535–570.
- Douven, Igor and Rainer Hegselmann. 2020. "Mis- and Disinformation in the Bounded Confidence Model." Submitted.
- Flache, Andreas, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. 2017. "Models of Social Influence: Towards the Next Frontiers." *Journal of Artificial Societies and Social Simulation* 20 (4).
- Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–1380.
- Harary, Frank, Robert Z. Norman, and Dorwin Cartwright. 1965. *Structural Models. An Introduction to the Theory of Directed Graphs*. New York: John Wiley & Sons, Inc.
- Hegselmann, Rainer and Ulrich Krause. 2002. "Opinion dynamics and bounded confidence: models, analysis and simulation." *Journal of Artificial Societies and Social Simulation* 5 (3).
- Hegselmann, Rainer and Ulrich Krause. 2015. "Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model." *Networks & Heterogeneous Media* 10 (3): 477–509.
- Hegselmann, Rainer, Stefan König, Sascha Kurz, Christoph Niemann, and Jörg Rambau. 2015. "Optimal Opinion Control: The Campaign Problem." *Journal of Artificial Societies and Social Simulation* 18 (3).

- Krause, Ulrich. 1997. "Soziale Dynamiken mit vielen Akteuren. Eine Problemskizze." Pp. 37–51 in *Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren*. Edited by Ulrich Krause and Manfred Stöckler. Bremen: Universität Bremen.
- Krause, Ulrich. 2000. "A discrete nonlinear and non-autonomous model of consensus formation." Pp. 227–36 in *Communications in Difference Equations. Proceedings of the Fourth International Conference on Difference Equations, Poznan, Poland, August 27–31, 1998*. Edited by Saber Elaydi, Gerry Ladas, Jerzy Popenda, and Jerzy Rakowski. Amsterdam: Gordon and Breach Science Publishers.
- Lehrer, Keith and Carl Wagner. 1981. *Rational consensus in science and society. A philosophical and mathematical study*. Dordrecht: D. Reidel Publishing Company.
- Lorenz, Jan. 2007. "Continuous opinion dynamics under bounded confidence: a survey." *International Journal of Modern Physics C* 18 (12): 1819–1838.
- Pariser, Eli. 2011. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin Press.
- Sirbu, Alina, Vittorio Loreto, and Vito D. P. Servedio. 2017. "Opinion Dynamics: Models, Extensions and External Effects." Pp. 363–401 in *Participatory Sensing, Opinions and Collective Awareness*. Edited by Vittorio Loreto, Muki Haklay, Andreas Hotho, Vito D.P. Servedio, Gerd Stumme, Jan Theunis, and Francesca Tria. Cham: Springer International Publishing.
- Sunstein, Cass R. 2017. *#republic: Divided Democracy in the Age of Social Media*. Princeton, N.J.: Princeton University Press, 2017.
- Urbig, Diemo, Jan Lorenz, and Heiko Herzberg. 2008. "Opinion Dynamics: the Effect of the Number of Peers Met at Once." *Journal of Artificial Societies and Social Simulation*, 11 (2).
- Wedin, Edwin and Hegarty, Peter. 2015. "The Hegselmann-Krause Dynamics for the Continuous-Agent Model and a Regular Opinion Function Do Not Always Lead to Consensus." *IEEE Transactions on Automatic Control* 60 (9): 2416–421.
- Xia, Haoxiang, Huili Wang, and Zhaoguo Xuan. 2011. "Opinion Dynamics: A Multidisciplinary Review and Perspective on Future Research." *International Journal of Knowledge and Systems Science* 2 (4): 72–91.

Advances in the Sociology of Trust and Cooperation

Theory, experiments, and field studies

Edited by

Vincent Buskens, Rense Corten and Chris Snijders

DE GRUYTER

ISBN 978-3-11-064701-3

e-ISBN (PDF) 978-3-11-064749-5

e-ISBN (EPUB) 978-3-11-064761-7

DOI <https://doi.org/10.1515/9783110647495>



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. For details go to: <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Library of Congress Control Number: 2020946148

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

©2020 Vincent Buskens, Rense Corten and Chris Snijders, published by Walter de Gruyter GmbH, Berlin/Boston

Cover image: Viesinsh/iStock/Getty Images Plus

Typesetting: Integra Software Services Pvt. Ltd.

Printing and Binding: CPI books GmbH, Leck

www.degruyter.com

Contents

Vincent Buskens, Rense Corten and Chris Snijders

- 1 Complementary Studies on Trust and Cooperation in Social Settings:
An Introduction — 1**

Part I: Theoretical Contributions

Thomas Voss

- 2 Institutional Design and Human Motivation: The Role of Homo
Economicus Assumptions — 15**

Karl-Dieter Opp

- 3 Rational Choice Theory, the Model of Frame Selection and Other
Dual-Process Theories. A Critical Comparison — 41**

Marcel A.L.M. van Assen and Jacob Dijkstra

- 4 Too Simple Models in Sociology: The Case of Exchange — 75**

Andreas Flache

- 5 Rational Exploitation of the Core by the Periphery? On the Collective
(In)efficiency of Endogenous Enforcement of Universal Conditional
Cooperation in a Core-Periphery Network — 91**

Siegwart Lindenberg, Rafael Wittek and Francesca Giardini

- 6 Reputation Effects, Embeddedness, and Granovetter's Error — 113**

Arnout van de Rijt and Vincenz Frey

- 7 Robustness of Reputation Cascades — 141**

Henk Flap and Wout Ultee

- 8 Organized Distrust: If it is there and that Effective, Why Three Recent
Scandals? — 153**

Rainer Hegselmann

- 9 Polarization and Radicalization in the Bounded Confidence Model:
A Computer-Aided Speculation — 199**

Michał Bojanowski

10 Local Brokerage Positions and Access to Unique Information — 229

Thomas Gautschi

11 Who Gets How Much in Which Relation? A Flexible Theory of Profit Splits in Networks and its Application to Complex Structures — 249

Part II: Experimental Tests

Ozan Aksoy

12 Social Identity and Social Value Orientations — 295

Fabian Winter and Andreas Diekmann

13 Does Money Change Everything? Priming Experiments in Situations of Strategic Interaction — 309

Martin Abraham, Kerstin Lorek and Bernhard Prosch

14 Social Norms and Commitments in Cooperatives – Experimental Evidence — 319

Hartmut Esser

15 Rational Choice or Framing? Two Approaches to Explain the Patterns in the Fehr-Gächter-Experiments on Cooperation and Punishment in the Contribution to Public Goods — 335

Christoph Engel and Axel Ockenfels

16 Maverick: Experimentally Testing a Conjecture of the Antitrust Authorities — 357

Rense Corten, Vincent Buskens and Stephanie Rosenkranz

17 Cooperation, Reputation Effects, and Network Dynamics: Experimental Evidence — 391

Davide Barrera, Vincent Buskens and Vera de Rover

18 Comparing Consequences of Carrots and Sticks on Cooperation in Repeated Public Good Games — 417

Part III: Field Studies

Gerrit Rooks, Chris Snijders and Frits Tazelaar

- 19 A Sociological View on Hierarchical Failure: The Effect of Organizational Rules on Exchange Performance in Buyer-Supplier Transactions — 443**

Ferry Koster

- 20 Organizational Innovativeness Through Inter-Organizational Ties — 465**

Anne Roeters, Esther de Ruijter and Tanja van der Lippe

- 21 A Transaction Cost Approach to Informal Care — 483**

Beate Volker

- 22 Trust is Good – Or is Control Better? Trust and Informal Control in Dutch Neighborhoods – Their Association and Consequences — 499**

Tom A.B. Snijders and Frank Kalter

- 23 Religious Diversity and Social Cohesion in German Classrooms: A Micro-Macro Study Based on Empirical Simulations — 525**

Notes on the Editors and Contributors — 545