The Relative Disagreement Model of Opinion Dynamics: Where Do Extremists Come From?

Michael Meadows, Dave Cliff

The Relative Disagreement model of opinion dynamics: where do extremists come from?

Michael Meadows and Dave Cliff[1]

Abstract

In this paper we introduce a novel model that can account for the spread of extreme opinions in a human population as a purely local, self-organising process. Our starting point is the well-known and influential Relative Agreement (RA) model of opinion dynamics introduced by Deffuant *et al.* (2002). The RA model explores the dynamics of opinions in populations that are initially seeded with some number of "extremist" individuals, who hold opinions at the far ends of a continuous spectrum of opinions represented in the abstract RA model as a real value in the range [-1.0, +1.0]; but where the majority of the individuals in the population are, at the outset, "moderates", holding opinions closer to the central mid-range value of 0.0. Various researchers have demonstrated that the RA model generates opinion dynamics in which the influence of the extremists on the moderates leads, over time, to the distribution of opinion values in the population converging to attractor states that can be qualitatively characterised as one of either uni-polar and bi-polar extremes, or reversion to the centre ("central convergence"). However, a major weakness of the RA model is that it pre-supposes the existence of extremist individuals, and hence says nothing to answer the question of "where do extremists come from?" In this paper, we introduce the Relative Disagreement (RD) model, in which extremist individual arise spontaneously and can then exert influence over moderates, forming large groups of polar extremists, via an entirely internal, self-organisation process. We demonstrate that the RD model can readily exhibit the uni-polar, bi-polar, and central-convergence attractors that characterise the dynamics of the RA model, and hence this is the first paper to describe an opinion dynamic model in which extremist positions can spontaneously arise and spread in a population via a self-organising process where opinion-influencing interactions between any two individuals are characterised not only by the extent to which they agree, but also by the extent to which they disagree.

IWSOS Key Topics: models of self-organisation in society; techniques and tools for modelling self-organising systems; self-organisation in complex networks; self-organising group and pattern formation; social aspects of self-organisation.

1. Introduction: Opinion Dynamics and the RA Model

Since its inception, "opinion dynamics" has come to refer to a broad class of different models applicable to many fields ranging from sociological phenomena to ethology and physics (Lorenz 2007). The focus of this paper is on an improvement of Deffuant *et al.*'s (2002) "Relative Agreement" (RA) model, that

[1] Department of Computer Science, University of Bristol, Bristol BS8 1UB, U.K.
Emails: michaeljmeadows86@gmail.com and dc@cs.bris.ac.uk.

was originally developed to assess the dynamics of political, religious and ideological extremist opinions, and the circumstances under which those opinions can rise to dominance via processes of self-organisation (i.e., purely by local interactions among members of a population) rather than via exogenous influence (i.e. where the opinion of each member of a population is influenced directly by an external factor, such as mass-media propaganda). The RA model was developed with the aim of helping to explain and understand the growth of extremism in human populations, an issue of particular societal relevance in recent decades where extremists of various religious or political beliefs have been linked with significant terrorist acts.

Suppose a group of *n* experts are tasked with reaching an agreement on a given subject. Initially, all the experts will possess an opinion that for simplicity we imagine can be represented as a real number *x*, marking a point on some continuum. During the course of their meeting, the experts present their opinion to the group in turn and then modify their own opinion in light of the views of the others, by some fixed weight. If all opinions are equal after the interaction, it can be said that a consensus has been reached, otherwise another round is required. It was demonstrated by de Groot (1974) that this simple model would always reach a consensus for any positive weight. Although highly abstract and clearly not particularly realistic, this simple model has become the basis for further analysis and subsequent models (e.g. Chatterjee & Seneta 1977; Friedkin 1999).

Building on the de Groot model, the Hegselmann-Krause model included the additional constraint that the experts would only consider the opinions of others that were not too dissimilar from their own (Krause 2000); this is also known as the Bounded Confidence (BC) model. The BC model adds the idea that each expert has a quantifiable conviction about their opinion, their uncertainty, *u*. It was demonstrated that although a consensus may be reached in the BC model, it is not guaranteed (Hegselmann & Krause 2002). It was observed that when the BC model is set in motion with every agent having an initially high confidence (low uncertainty) about their own random opinion, the population disaggregates into large numbers of small clusters; and as the uncertainty was increased, so the dynamics of the model tended towards those of the original de Groot model (Krause 2000). Later, the model was tested with the inclusion of "extremist" agents, defined as individuals having extreme value opinions and very low uncertainties. In the presence of extremists it was found that the population could tend towards two main outcomes: *central convergence* and *bipolar convergence* (Hegselmann & Krause 2002). In central convergence, typical when uncertainties are low, the majority of the population clustered around the central, "moderate" opinion. In contrast, when uncertainties were initially high, the moderate population would split into two approximately equal groups one of which would tend towards the positive extreme and the other towards the negative: referred to as *bipolar convergence*.

Although these two phenomena have occurred in real human societies, there is a third well-known phenomenon that the BC model is unable to exhibit: an initially moderate population tending towards a single extreme (and hence known as *single extreme convergence*).

Shortly after the publication of the BC model, Deffuant, Amblard, Weisbuch, & Faure (2002) reported their exploration of the BC model and proposed an extension of it which they named the Relative Agreement (RA) model (Deffuant *et al.* 2002). The RA model was intended to be capable of exhibiting single extreme convergence in its dynamics.

There are two main differences between the RA model and the BC model. The first change is that agents are no longer expressing their opinion to the group as a whole followed by a group-wide opinion update. Instead, in the RA model pairs of agents are randomly chosen to interact and update. This is repeated until stable clusters have formed. The second change relates to how agents update their opinions. In the BC model an agent only accepted an opinion if it fell within the bounds of their own uncertainty, and the weight that was applied to that opinion was fixed. In the RA model however, an opinion is weighted proportional to the degree of overlap between the uncertainties of the two interacting agents.

These changes represent a push for increased realism. In large populations, individuals cannot necessarily consider the opinion of every other agent; therefore paired interactions are far more plausible. More importantly, the RA model also allows for agents with strong convictions to be far more convincing than those who are uncertain (Deffuant 2006). Thus, although the RA model is stochastic, the only random element of the model is in the selection of the individuals for the paired interactions (Lorenz 2005). As expected, the RA model was able to almost completely replicate the key results of the BC model (Deffuant *et al.* 2000).

Having demonstrated that RA model was comparable to the BC model under normal circumsthances, Deffuant *et al.* then added the *extremist* agents to the population, to see if they could cause shifts in the opinions of entire population. An extremist was defined as an agent with an opinion above 0.8 or below -0.8 and with a very low uncertainty. Conversely, a moderate agent is one whose absolute opinion value is less than 0.8 and with a fixed, higher uncertainty who is therefore more willing to be persuaded by other agents. Under these circumstances, Deffuant *et al.* reported that there are large areas of parameter space in which all three main types of population convergence could occur. The fact that the RA model offers realistic parameter-settings under which single extreme convergence regularly occurs is a particularly novel attraction.

To classify population convergences, Deffuant *et al.* (2002) introduced the y metric, defined as: $y = p'^{2}_{+} + p'^{2}_{-}$ where p'_{+} and p'_{-} are the proportion of initially moderate agents that have finished with an opinion that is classified as extreme at the positive and negative end of the scale respectively. Thus, central, bipolar and single extreme convergences have y values of 0.0, 0.5 and 1.0, respectively.

Meadows & Cliff (2012) recently demonstrated that while all three population convergences can indeed occur using various parameter settings, the specific parameter values that do allow for the more extreme convergences are not as originally reported by Deffuant *et al.* (2002). Meadows & Cliff state that they reviewed over 150 papers that cite Deffuant *et al.* (2002) but not a single one of them involved an independent replication of the original published RA results. In attempting to then replicate the results reported by Deffuant *et al.* (2002), Meadows & Cliff (2012) found a significantly different, but simpler and more intuitively appealing result, as is illustrated here in Figures 1 and 2. In Figure 2, as agents' initial uncertainty *u* increases, the instability of the population rises with it, resulting in a higher *y* value; also, as the proportion of initially extremist agents increases, there is again a corresponding rise in the resultant instability. When there is a higher level of instability in the population, Figure 2 shows that there is a greater chance of the population converging in a manner other than simply to the centre. Also, in Figure 2 (unlike Figure 1) there is no area that implies a *guaranteed* single extreme convergence, although there is a large area of parameter space in which that can occur. This makes intuitive sense because a population with an initial instability that would allow single extreme convergence must surely also be unstable enough to allow central and bipolar convergences.

2. The Relative Disagreement Model

While the ability to produce a single extreme is significant, it should be noted that the RA model requires a significant minority (20-30%) of extremists be seeded into the initial population. Although useful as an academic tool, it could be argued that this is a somewhat artificial solution as it simply raises further questions, most notably, "Where did all of those extremist agents come from in the first place?" A model that could exhibit similar results to the RA model without the need for such a high proportion of extremist agents would be seen as a major improvement. One that could do away with the extremist agents altogether would be a considerable leap forward. The remainder of this paper will focus on the specification and examination of a model, quite like the RA model, that achieves similar results without the need for initially extreme agents.

The crux of this improvement lies in the observation that the RA model focuses only on the behaviour of agents when they are in agreement, yet it has long been known to psychologists that disagreements in opinions can lead to the distancing of the two opinions. That is to say, if you are debating with someone whose opinion is completely different to your own, you are likely to "update" your own opinion even further away from that with which you disagree. This is called *reactance* (Baron & Byrne 1991) and can be thought of as analogous to the more general "Boomerang effect" (Brehm & Brehm 1981) where someone attempting to influence an individual inadvertently produces the opposite effect in attitude change. Given this additional real-world information, there follows a formal definition of the proposed "relative disagreement" (RD) model; after the formal definition, we present an empirical exploration of its opinion dynamics.

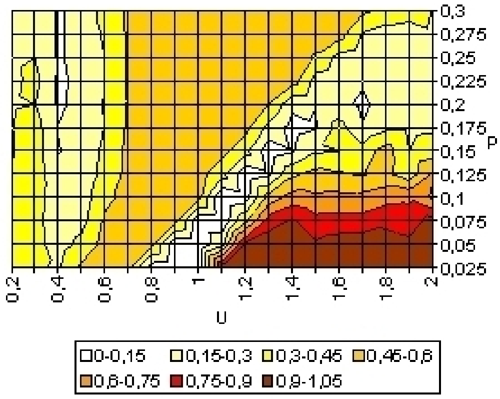


Figure 1: Average y values over the (p_e, u) parameter-space, reproduced directly from Deffuant *et al.* (2002), Figure 9.

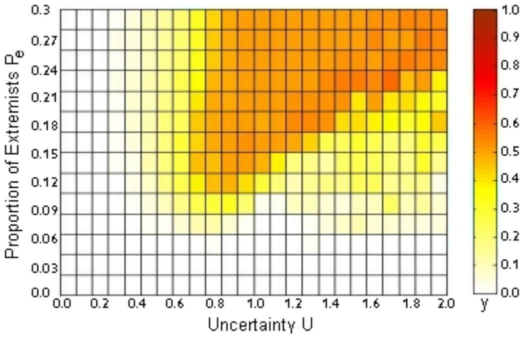


Figure 2: Average y values over the (p_e, u) parameter-space, as reported by Meadows & Cliff (2012); the zone of parameter space, and the values of other RA parameters, make this plot a direct correlate of the original shown in Figure 1. Reproduced directly from Meadows & Cliff (2012) Figure 12.

If we return to our population of n agents, each individual i is in possession of two variables; an opinion x, and an uncertainty u, both of which are represented by real numbers. In the RA model, the opinion could be initially set on the range of -1.0 to 1.0, with extremists being defined as agents whose opinions lay below -0.8 or above 0.8. As our goal is to replicate the convergences on the RA model without extremist agents we do not allow an opinion to be initially set outside the range of -0.8 and 0.8, but we retain the minimal and maximal values as before. Since we have no extremist agents, we are no longer constrained by defining agents by their opinions and so uncertainties are assigned randomly using a simple method to bias agents towards being uncertain (as it is in uncertain populations that more interesting results are to be found) given by:

$$u = min(\ random(0.2,\ 2.0) + random(0.0,\ 1.0),\ 2.0)$$

Random paired interactions take place between agents until a stable opinion state is produced. The relative agreement between agents *i* and *j* is calculated as before by taking the overlap between the two agents' bounds h_{ij}, given by:

$$h_{ij} = min\ (x_i + u_i,\ x_j + u_j) - max(x_i - u_i,\ x_j - u_j)$$

Followed by subtracting the size of the non-overlapping part given by:

$$2u_i - h_{ij}$$

So the total agreement between the two agents is given as:

$$h_{ij} - (2u_i - h_{ij}) = 2(h_{ij} - u_i)$$

Once that is calculated, the relative agreement is then given by:

$$2(h_{ij} - u_i)\ /\ 2u_i = (h_{ij}\ /\ u_i) - 1$$

Then if $h_{ij} > u_i$, then update of x_j and u_j is given by:

$$x_j := x_{\underline{j}} + \mu_{RA}[(h_{ij}\ /\ u_i) - 1](x_i - x_j)$$
$$u_j := u_{\underline{j}} + \mu_{RA}[(h_{ij}\ /\ u_i) - 1](u_i - u_j)$$

Similarly, the relative disagreement between agents *i* and *j* is calculated by a very similar method to find g_{ij}:

$$g_{ij} = max(x_i - u_i,\ x_j - u_j) - min\ (x_i + u_i,\ x_j + u_j)$$

Followed by subtracting the size of the non-overlapping part given by:

$$2u_i - g_{ij}$$

So the total disagreement between the two agents is given as:

$$g_{ij} - (2u_i - g_{ij}) = 2(g_{ij} - u_i)$$

Once that is calculated, the relative disagreement is then given by:

$$2(g_{ij} - u_i)\ /\ 2u_i = (g_{ij}\ /\ u_i) - 1$$

We have chosen to use an analogous method for calculating the agents' disagreement for ease of understanding as it also facilitates the need for calculating relative disagreement. Now we would not want the disagreement update to occur in every instance of disagreement, as it is intuitively obvious that this would not occur in the every real-world instance of disagreement. Therefore if $g_{ij} > u_i$ and with a probability λ, the update of x_j and u_j is given by:

$$x_j := x_j - \mu_{RD}[(g_{ij} / u_i) - 1](x_i - x_j)$$
$$u_j := u_j + \mu_{RD}[(g_{ij} / u_i) - 1](u_i - u_j)$$

Note that with this model, the definition of g_{ij} will be equal to that of h_{ij}, but it is important to express this definition in full for a greater intuitive understanding of the model and the logic behind its construction.

3. Results of the RD model.

3.1 Comparing the RA and RD model

As has been stated previously, the RD model can only be considered to be an improvement of the RA model if it can replicate the results of the RA model without the aid of initially extreme agents. This comparison however, leads to an initial problem that is worth noting now. When analysing the RA model we can examine individual sample runs (as shown in the next section) or typical patterns of *y* like those in Figure 1 and 2.

As shown those figures, we see that the main parameters examined when we look for various patterns of *y* are the initial uncertainty of the moderate agents (equivalent to all agents in the RD model) and the proportion of initially extreme agents in the population, which clearly presents us with a problem. Neither of those values are relevant to the RD model; uncertainties are randomised and there are no longer any extremist agents. Therefore the patterns of *y* that we shall find with the RD model will not be analogous with those from the RA model. Indeed, in terms of comparing the RA and RD models, the first and most convincing step will be to demonstrate that all three convergences are possible before we move on to looking at new methods of examining *y*-values.

3.2 Reproducing convergences

Clearly the most basic function of any model aiming to be comparable to the RA model is to be able to demonstrate the three main population opinion convergences: central, bipolar and single extreme. In addition, a successful model should be able to perform these convergences without any drastic alterations to the parameters in the original RA model, much like the specification provided in Section 2. As would be expected, we find that in this model a population converges towards the centre in a majority of the simulation runs, more so than in the RA model. We believe that this is understandable because we no longer have 30% of the agents functioning solely to skew the overall opinion of the population towards their own. Any aberration in the population's opinion must come from agreements and importantly, disagreements from within the body of moderate agents. Nonetheless, we found that central and bipolar convergences can occur regularly in given parameter spaces and that a single extreme convergence can occur when the uncertainties are randomised as prescribed previously as shown in Figures 3, 4, and 5.

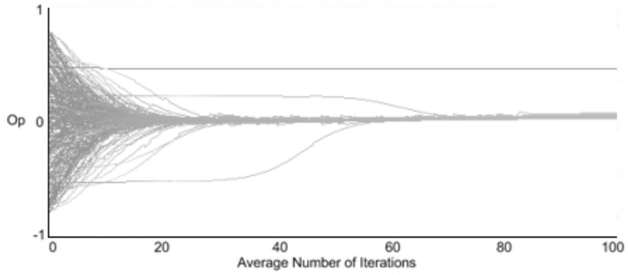
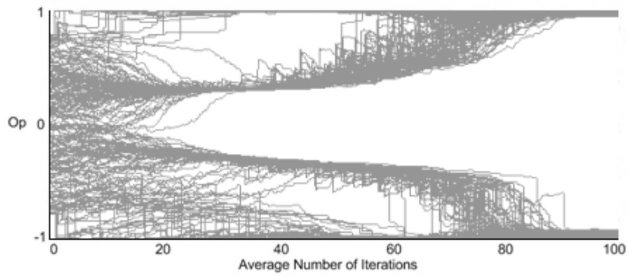


Figure 3: An example of central convergence in the RD model with n = 200, U = 0.8, μ_{RA} = 0.05, μ_{RD} = 0.05, λ = 0. 0.



.

Figure 4: An example of bipolar convergence in the RD model with n = 200, U = 0.4, μ_{RA} = 0.2, μ_{RD} = 0.2, λ = 0.1.

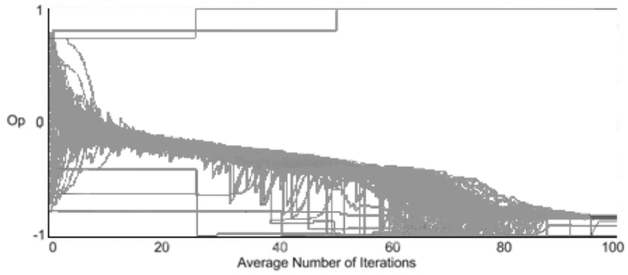


Figure 5: An example of single extreme convergence in the RD model with n = 200, U = 0.1, μ_{RA} = 0.5, μ_{RD} = 0.7, λ = 0.4.

3.3 Examining new patterns of y

Having already established that the heat-map graphs visualising how the y-value changes with respect to p_e (the proportion of initially extreme agents) and u are no longer relevant, we need to employ a new visualisation technique for illustrating patterns of y-values. It is beyond the scope of this paper to examine the whole variety of different options available, and so here we will focus solely on an examination of how the typical y-value changes as we alter the values μ_{RA} and μ_{RD} (the degree to which agents affect each other when they agree and disagree respectively) with different values of λ (the probability that our disagreements result in an action), as shown in Figure 6.

It is immediately clear from Figure 6 that the RD model provides much higher population stability than the RA model. Note that almost every y-value heat-map is almost entirely white, indicating a very high proportion of simulation runs that resulted in central convergence. This is of course not a disappointing result and should be expected; we no longer have extremist agents manipulating our moderate agents. It therefore becomes more interesting to examine the various standard deviation heat-map graphs.

In the top row of Figure 6, we see that there is nothing particularly special happening. This is because we have $\lambda = 0.0$, and thus the model behaves exactly as the RA model would without any extremist agents and as such, we see a consistent central convergence with no variation.

Once λ is increased, we see an immediate increase in population instability; and effect analogous to increasing the initial uncertainty u, or the proportion of extremist agents, in the RA model. It is in these areas of instability that we find the possibility of a population converging in a bipolar or single extreme manner although, as with the RA model, these instabilities do not guarantee that bipolar or single extreme convergences *will* occur, merely that they *could* occur. That is: there is an increased probability of such instabilities occurring but the probability is not unity. Not only is this encouraging in terms of validating the RD model with respect to the RA model findings given by Meadows & Cliff (2012), but also in terms of realism: in the majority of model runs, the population remain moderate with bipolar and single extreme populations being a small minority of outcomes.

A final point of interest from Figure 6 would be to observe that the increase in λ significantly increases the size and level of instability inherent to the population. This is of course to be expected, because as we see in any one particular graph, as μ_{RD} increases, the more unstable the population becomes, implying that increasing the influence that a disagreement can have will increase the population. It also explains the observation that as λ increases, the need for μ_{RD} to be a high value to allow for population instability decreases quite substantially.

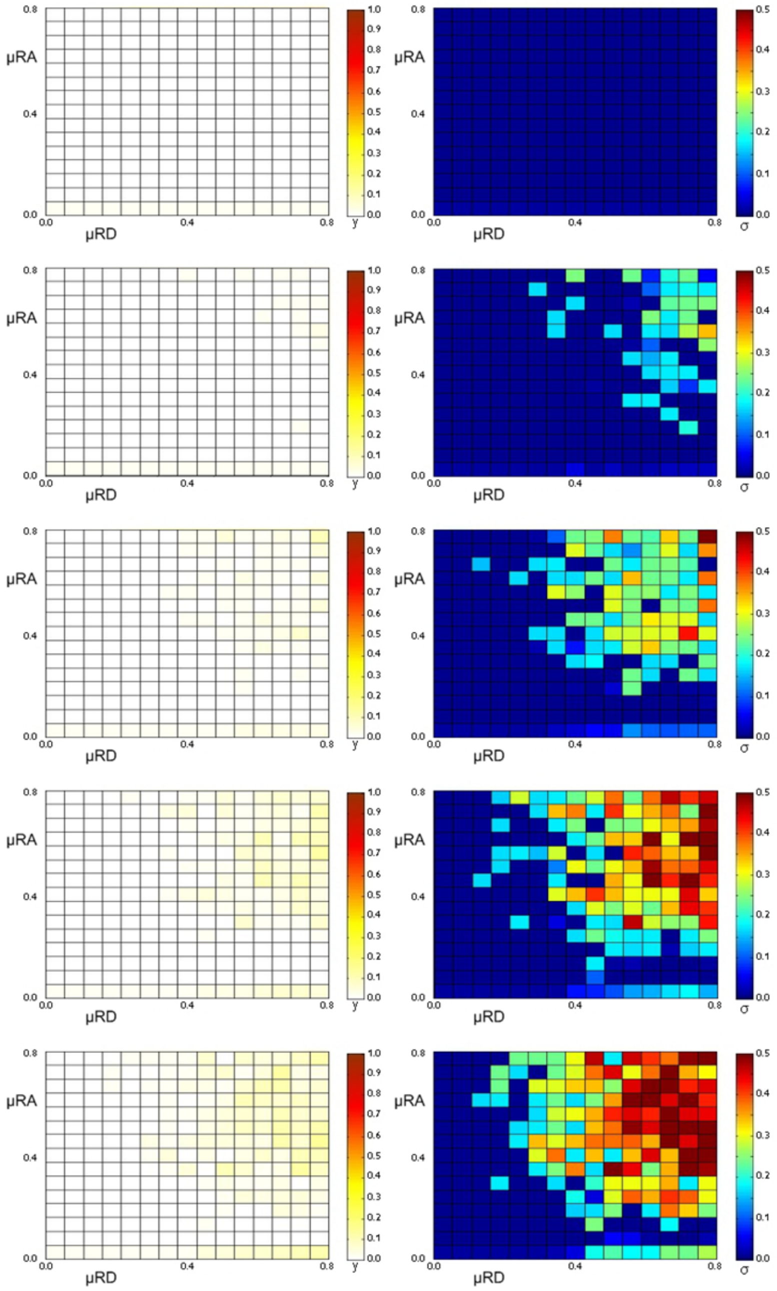


Figure 6: Average y values (left column) and standard deviations (right column) for (from top to bottom) $\lambda = 0.0$, 0.25, 0.5, 0.75, and 1.0.

4. Further work

As has been shown, this examination of the RD model indicates that the introduction of extremist agents in an otherwise moderate population is not the only way to create a population-wide instability. By removing the extremists and allowing the moderate agents to generate the instability themselves, via a self-organising process, we see an increase in realism while answering the question of where the significant RA extremist minority would have come from in the first place. One potential line of further research would be to introduce clustered populations with nontrivial topologies in the "social network" of who can influence who, as explored by Amblard & Deffuant (2004) who experimented with small-world networks; and more recently by [Author names deleted for blind review] (2013) who extended the study of RA networks to look at opinion dynamics on social networks with scale-free and small-world topologies, and also on networks with "hybrid" scale-free/small-world topological characteristics as generated by the Klemm-Eguiluz (2002) network construction algorithm. Exploration of the dynamics of the RD model on such complex networks would allow for yet another push towards realism in the RD model whilst at the same time allowing for further comparison of results results against the RA model to ensure applicability. Once that has been completed it would be worth investigating other application areas for the RD model. At the moment it is clear that the RD model could be applied to much larger fields outside of extremist opinions and behaviour relating to terrorist nature (for example, aiding in business marketing strategies and techniques).

5. Conclusion

In this paper we have presented, for the first time, the RD model. It is clear that the RD model represents a significant advance in the understanding of opinion dynamics and can be considered a logical next step after the RA model. It is also clear that although the properties of this new system are beginning to be uncovered, there is already a great deal of scope for adding extra realism to the model.

References

AMBLARD, F., & Deffuant, G. (2004) The role of network topology on extremism propagation with the Relative Agreement opinion dynamics. *Physica A*. **343:** 725-738.

[Author names deleted for blind review] (2013) The Relative Agreement model of opinion dynamics in populations with complex social network structure. Submiited to *CompleNet2013: the Fourth International Workshop on Complex Networks.* Berlin, March 2013.

BARON, R., & Byrne, D. (1991). *Social Psychology*. 6th Edition. Boston: Allyn and Bacon.

BREHM, S. & Brehm, J. (1981) *Psychological reactance: a theory of freedom and control.* New York: Academic Press.

CHATTERJEE, S. & Seneta, E. (1977) Towards Consensus: Some Convergence Theorems on Repeated Averaging. *Journal of Applied Probability,* **14**(1): 88-97.

DE GROOT, M. (1974) Reaching a Consensus. *Journal of the American Statistical Association,* **69**(345): 118-121.

DEFFUANT, G., Neau, D., & Amblard, F. (2000) Mixing beliefs among interacting agents. *Advances in Complex Systems,* **3**: 87-98.DEFFUANT, G., Amblard, F., Weisbuch, G. and Faure, T. (2002) How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation,* **5**(4):1.

DEFFUANT, G. (2006) Comparing Extremism Propagation Patterns in Continuous Opinion Models. *Journal of Artificial Societies and Social Simulation,* **9**(3): 8.

FRIEDKIN, N. (1999) Choice Shift and Group Polarization. *American Sociological Review,* **64**(6): 856-875.

HEGSELMANN, R. & Krause, U. (2002) Opinion dynamics and bounded confidence: models, analysis and simulation. *Journal of Artificial Societies and Social Simulation,* **5**(3): 2.

KLEMM, K. & Eguíluz, V. (2002) Growing Scale-Free Networks with Small-World Behavior. *Physical Review E,* **65**: 057102.

KOZMA, B. & Barrat, A. (2008) Consensus formation on adaptive networks. *Physical Review E,* **77**:016102

KRAUSE U. (2000). A Discrete Nonlinear and Non-Autonomous Model of ConsensusFormation. *Communications in difference equations: proceedings of the Fourth International Conference on Difference Equations, August 27-31, 1998*, 227-236.

LORENZ, J. & Deffuant, G. (2005) The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Physica A: Statistical Mechanics and its Applications,* **343**: 725-738.

LORENZ, J. (2007) Continuous Opinion Dynamics under Bounded Confidence: A Survey. *International Journal of Modern Physics C* **18***:*1-20.

MEADOWS, M. & Cliff, D. (2012) Reexamining the Relative Agreement Model of Opinion Dynamics. *Journal of Artificial Societies and Social Simulation,* **15**(4):4.

SOBKOWICZ, P. (2009) Studies on opinion stability for small dynamic networks with opportunistic agents. *International Journal of Modern Physics C*, **20**(10) 1645-1662.