The grand challenge of *helping people agree* and how we might go about collectively tackling it

Bruce Edmonds[1][0000-0002-3903-2507], Dino Carpentras[2][0000-0001-8471-2352] and Edmund Chattoe-Brown[3][0000-0001-8232-6896]

[1]Centre for Policy Modelling, Manchester Metropolitan University, UK
bruce@edmonds.name
[2]Department of Humanities, Social and Political Science, ETH Zurich, CH
dino.carpentras@gess.ethz.ch
[3]School of Media Communication and Sociology, University of Leicester, UK
ecb18@leicester.ac.uk

Abstract. Despite the fact that the world needs agreement in order to address threats, such as climate change, people find this hard, even when all concerned desire an agreement. Social simulations have explored a number of different mechanisms relevant to such processes, including those about: opinion dynamics, negotiation, social identity, collective intelligence and voting models. However these models tend to: (1) stick to their own silos so the connections between strands are only explored sporadically, (2) not engage sufficiently with real-world cases/data to be useful, and (3) avoid modelling transitions between different "modes" of interaction. We sketch some elements of a programme to bring these kinds of modelling together to address this grand challenge. We call upon the social simulation community to coordinate, with the ultimate goal of finding ways to facilitate such agreement processes, before it is too late.

Keywords: agreement, dialogue, polarisation, negotiation, social influence, voting, empirical support, phase transition.

1 The Challenge

The world faces a variety of global challenges, including: war, climate change, water shortages, species loss, epidemics and poverty. There are similar challenges at national, regional and local scales. In order to meet these challenges effectively, people will need to coordinate their actions, coming to agreement as to joint action plans [16]. However, there is a major problem – people find this very hard to do in practice. Polarisation, group identity, mutual incomprehension, differing world views, fake news, ignorance and sheer lack of communication can prevent progress towards agreement, *even* if this is an outcome everyone involved wants. This does not address the situation where the parties do not *want* to agree, but that still leaves many cases where agreement is desired.

A typical case is where people somehow see members of another group as enemies, as this perception renders agreement a "zero sum game." However, the perspective that creates this situation is not immutable. In fact, in the United States, affective polarisation (i.e. dislike for members of the other party) has been rising over the years [2]. This implies that change is possible and that maybe the process can be reversed, recreating circumstances where agreement and win-win situations are again possible despite differing interests of members in each party. It would be really useful if we developed a set of guidelines and interventions that helped maintain higher levels of useful dialogue, even when interests are divergent, but equally might help us de-escalate highly polarised situations to reduce ineffective confrontation. We know that these transitions do occur (e.g. in the Good Friday Agreement), but our understanding of them is weak. Thus the "grand challenge" that we present to the social simulation community is this:

Can social simulation inform us as how to manage situations to facilitate agreement between parties when this is desired or necessary?

The challenge is an ambitious one. There are many different mechanisms facilitating or frustrating potential agreement and many aspects of a situation that might need to be taken into account (only some of which may be recognised by existing families of models). Coming to a collective agreement involves cognitive, social and institutional processes, which makes it hard for fields focusing only on the micro (e.g. psychology) or the macro (e.g. quantitative sociology). Such interactive processes are inherently dynamic so understanding them by only considering snapshots of evidence (e.g. single surveys) is hard. There are texts on the *art* of helping groups to agreement, but without formal models, these are not precise as to when a particular structure, technique or intervention will be helpful. Social simulations are a perfect candidate as they can combine the micro and macro aspects in a dynamic way [11].

However, the challenge also needs to be feasible – something the social simulation community could attempt. We are not going to be able to address all the difficulties of working democracies and we are not going to be able to directly influence what individuals believe at the start of a process. What we can perhaps do is offer suggestions in situations where a group of people or representatives that (in general terms) seek agreement, struggle with a variety of goals, beliefs, cultures, and relationships amongst the participants. We might be able to identify: some of the difficulties (which, in keeping with the strengths of Agent-Based Modelling (ABM), may be processual and counter-intuitive) and some of the ways to increase the chance or extent of agreement.

This paper does not present a unified model for facilitating agreement, as we are still quite far from this goal. Instead, we discuss how we can collectively develop a more comprehensive, systematic and coherent approach.

2 Some existing modelling work

One strategy to open up this challenge is to look at the ideas about conflict and agreement already encapsulated in models. There are many bodies of social simulation work that are relevant to understanding how the opinions or beliefs of sets of interacting individuals might develop. On the whole, each of these has remained separate from the others developing its own distinct challenges, conventions and modelling approaches. We briefly mention some of these bodies of work (this is not a complete list), drawing out a few examples of what might be helpful with respect to the challenge and subsequently considering more general issues that arise from comparison between approaches. The point here is to give an idea of the diversity between approaches and methodologies – the different kinds of mechanism they cover – but also point out each one's limitations and assumptions. We touch upon: opinion/cultural dynamics, negotiation, social identity, social norms, and collective intelligence.

Opinion dynamics (OD) is a field of research exploring mechanisms and processes behind the evolution of opinions in society [14], focussing on how people are influenced by and influence other people's opinions. Despite this general goal, OD models differ in a variety of aspects. In models using attractive forces agents always become more similar when they interact (e.g. [10]). A special case of this are the "bounded confidence models," where agents cannot interact if their opinions are too different (e.g. [9]). In models with repulsive forces, agents can push each other towards opposite extremes if their opinions differ enough (e.g. [20]). In reinforcement systems, agents who agree push each other in the same direction (e.g. [23]). Other important distinctions include how opinions are represented within the agents, with "Cultural Dynamics" usually referring to models which employ multi-dimensional categorical opinions, while models that fall under the category of "sociophysics" usually employ unidimensional binary opinions [5]. Most of the models are based on the idea of "social influence," – that people become more similar when they interact [14], but this is not a specific theory. Some authors ground their model on more specific psychological theories and phenomena, such as "social impact theory" (e.g. [18]) or cognitive dissonance (e.g. [15]), while others derive their model directly from experiments (e.g. [4]). OD models have been used to show how simple rules can produce macroscopic effects such as different cultural groups and polarisation. To date, many models and effects have shown the possibility of creating this diversity also from homogeneous initial conditions, however, much less is known of which effects are actually responsible for the formation and evolution of groups in the real world.

Negotiation. The essence of negotiation models is that, by reaching agreement, the negotiating parties can access outcomes that they could not achieve alone. Providing a negotiation concludes. Participants do not necessarily start with the same model of the situation or full knowledge of the other participants (indeed some may actively *misrepresent* their interests). Despite this, many models under the label of "negotiation" only represent haggling over continuous dimensions with the communication between actors limited to suggestions as to the price, quality etc. e.g. [19]. In these, there is no communication about any other actions nor about their wider goals or what might be possible in the situation as is observed [31]. To use an example from one of the relatively rare models of negotiation applying to human actors, [13]. A lot of negotiation in practice is not about the reality of what people can afford but about the expectations they have of others and those they create in others.

Social identity includes a wide range of mechanisms whereby perceptions of an individual's or other's identity affects how one behaves towards them. In some earlier models such groups were defined by emerging similar 'tags' (observable markers), e.g. [1]. Here a propensity to interact with those with similar tags results in the emergence of cooperative groups. However, it is also the fact that perceptions of groups as entities affect how people behave. These two aspects are integrated within the "Social Identity Approach" (SIA), which is an integration of the sociological theory of Social Identity Theory [28] – how perceived groupings affect people's behaviour (e.g. biased towards their in-group) – and Self-categorization Theory [29] which takes a more cognitive approach. It explains social differentiation and its consequences, by including descriptions of various social, material and political forces [24]. There is now some interest in terms of implementing the SIA approach in terms of simulation models, with a special issue in JASSS, including a review of such models [26]. Social identity is often salient but the question of when it is and is not salient, is a more tricky matter. Thus, whilst this set of ideas could be applied in the modelling of many discursive situations it is unclear when this is necessary as it can add considerable complexity.

Social norms are prescribed guides for conduct that agents infer from those around them, through explicit or implicit normative declarations such as "You should not drop litter" [30] and sometimes conveyed in the form of statements like "Smoking is antisocial behaviour" [7]. As with social identity this involves both cognitive and social elements – there must be a perceptible pattern or convention, but also the individuals must believe that this is a norm [8]. The interaction of these two levels can result in complex dynamics, with norms coming into being and falling into decay [32]. The most ambitious framework to include all this complexity within socio-cognitive simulation is [8], which has been extended in a number of ways, e.g. to include values [17]. Whilst norms often constrain the manner in which discussions take place, within published models of norms they do not also touch upon the contents of such discussions. Anecdotally, norms (e.g. about political debate) can have a big influence on how political discussions proceed, but this has not yet been formally modelled.

Collective intelligence is often defined as "groups of individuals acting collectively in ways that seem intelligent" [22]. In practice, this often means that a group of people can outperform its members. The term "wisdom of the crowds" generally refers specifically to the task of guessing the correct answer to a question [21]. Models of collective intelligence often involve agents exploring a multi-dimensional "design space" in which a utility function is defined. The utility function is often modelled using the NK model, which produces a "rugged landscape" with multiple local maxima [27]. Some research has shown that connectedness can harm collective intelligence by encouraging people to follow the current best solution, rather than exploring the design space for better solutions [6,21]. Conversely, diversity has been shown to be a driver of better exploration of the solution space [25]. Collective intelligence models differ in the way they model agent characteristics, agent interactions, and how the design space and utility function are formalised [27]. Typically, models assume that the utility function is identical for all agents, but in many situations this should not be the case. This offers an interesting connection to opinion dynamics models, as the personalised utility could be considered an agent's opinion about a specific state of the system. Due to its optimisation goal this can be hard to relate to models about agent agreement.

To summarise, each of the above sub-fields has its own history, norms and goals, so models rarely cross their boundaries by including more than one approach. Maybe as a result, many of these have difficulties in establishing strong empirical links with observed cases, and tend to remain more at the level of an helpful analogy.

3 Different kinds of mass interaction

The differing families of models might partially map onto the different social mechanisms that might be involved. Different kinds of mechanism seem to be differently involved in observed processes, including the following kinds of situations.

1. *Apathy* – most actors are not motivated to discuss the issues with each other, but might vaguely listen to a few opinion leaders/politicians and might occasionally vote
2. *Polarised groups* – when there are distinct groups in opposition to each other with no meaningful dialogue between groups (other than to annoy each other)
3. *Negotiation* – when different parties try to reach a negotiated solution, involving give-and-take, issue framing, mapping goals or areas of agreement etc.
4. *Connected influence* – where the issues/opinions are not so much evidentially rooted but are actively spread to people one knows in a decentralised process.
5. *Additive collective intelligence* – when there is a constructive process of adding knowledge together to reach better solutions than any of the individuals could.

It is important to understand that these situations are often intertwined. For example, "putting pineapple on pizza" and the correct pronunciation of "gif" are both memes and polarising topics. Also, they may transition to one of the other kinds, for example vaccination has transitioned in many countries from an "apathetic" to a polarised state, and facts, such as the shape of Earth are sometimes treated as opinions despite being facts. This suggests that it might be possible to combine all these different situations in a single model.

4 Specific research steps to tackle this challenge

Given the situation as sketched above, we suggest some of the steps that might be needed to meet this challenge. Basically, we are suggesting a classic "divide and conquer" approach to this, starting from where we are (the different families of models) and building up to a more complex and integrated understanding later.

Assess the evidential basis for each kind of model. For each family of models (such as those families outlined in Section 2), an assessment needs to be made as to what it tells us – what understanding comes out of each that has the potential to help us understand observed discursive processes. This could be quite abstract as in providing counter-examples to plausible assumptions or a more applied result in terms of success as supporting empirically-based explanations of some observed patterns. This assessment should not be from the point of view of the modeller, but from the view of someone trying to understand an observed case.

Map the conditions of application of each kind of model to different kinds of observed situations. The assessment of each family of models will help us map these onto the "kinds" of observed interaction outlined in Section 3. This will likely not be a simple one-one map nor will one model apply everywhere, but something more on the lines of lists of statements such as: "Model type X can help us understand Y about interaction kind Z". Any such mapping is useful, even if it is partial and has overlaps (e.g. two families of models might tell you something about interaction kind Z).

Identify the "gaps" where there seem to be a lack of models. Whilst a complete mapping from families of model to kinds of situation is probably infeasible, identifying "gaps" where there are currently no adequate models is important as it indicates a new modelling sub-project is needed.

Understand when transitions between different kinds of interaction occur and why. Whilst some of the kinds of situation described in Section 3 might be quite "stable" (once one has got into a polarised situation where the parties do not communicate effectively then it may be hard to change), sometimes it does seem that situations do change from one kind to another. Thus, both modelling and empirical work is needed to understand such transitions. This is, maybe, one of the most challenging but also important, parts of the programme we are suggesting.

Suggest and test ways of supporting movement towards agreement for each kind of interaction. For each kind of situation, the modelling and empirical research could inform us as to how to encourage agreement from there. This may be quite weak advice on the lines of "Avoid doing X" but also might be something like "If things get to X then you could try Y or Z". It is *very* important that any such advice should not be overstated but reflects the evidence and understanding that has been established.

Suggest and test ways of encouraging transitions to different "phases" where agreement is more likely. Sometimes a change to a more "productive" kind of interaction seems to be essential to getting to a situation where an agreement could be reached. This is a hard task, and there is unlikely to be any simple recipes we can offer, but anything that increases the chance of change could be helpful.

This is not a comprehensive list of specific tasks to be done to tackle this challenge, but a suggestion for a programme to structure such a collective effort. Indeed, we expect this list to evolve in time as more research effort is dedicated to this challenge.

5 General community efforts to tackle this challenge

In addition to the specific steps outlined in the previous section, the development of models and methods for facilitating agreement between parties would require a broader effort from the community. It is not enough to simply develop models that can capture the cognitive and social processes involved in collective decision-making; we must also change the way we approach social simulations research.

Interaction between different fields and the development of a common language and common abstractions. One of the crucial issues of research is that often it ends up producing specialised sectors that "do not interact." This is problematic for almost all fields, as it hinders the production of new ideas and wastes time as people reinvent what has already been developed by others. However, for the goal described in this article, the situation is even worse, as it strongly requires connecting bits of knowledge which are currently scattered over many different fields.

Unfortunately, the social simulations community experiences both separation from other fields, as well as some degree of internal compartmentalization. As an example, we could look at the opinion dynamics literature. Despite their implication for psychology and sociology, these models are often written for readership in ABM and are read and cited mostly by other people within the same field. Besides this division with other disciplines, opinion dynamics experiences also some internal divisions, for example, between cultural dynamics models and bounded confidence models. It is not clear if they are simply two different formalisations of the same phenomenon or if they present some core difference besides their mathematical formulation. Similarly, within opinion dynamics there is still no clarity on terms like "opinion," "validation," and "experiment." For example, those interested in purely theoretical models use the term "experiment" to mean a specific simulation parameterisation, while scholars combining ABM with empirical research usually mean experiments on human participants.

To achieve the discussed goal, we need to be able to move across multiple disciplines, speak a sufficiently common language and possibly develop some common formalisation and standards.

Data inclusion and robustness tests. One of the key advantages of modelling how people could agree, is that this could be used for designing better policies and government structures. However, to achieve this goal, we should be sure about the validity of the model, otherwise we might be giving erroneous advice that makes policy failures more likely. A first way to check for output quality is to check its robustness. Two common methodologies are "sensitivity analysis," which checks the relationship between parameters and model's output, and robustness to noise. However, it is important to notice that the formalisation of the model is likely not unique (i.e. different researchers will produce different models of the same phenomenon/idea), and that also this type of robustness should be checked. There is evidence that in many cases ABMs can produce very different outputs due to imperceptibly small changes in its specification [12]. Similarly, it has been shown that the dynamics of the model can be completely altered just by choosing a different measurement scale [3]. This makes it very complex to distinguish between artefacts and correct model outputs.

Due to these limitations, in the last couple of decades, there have been multiple calls for data inclusion in fields such as opinion dynamics [14]. Despite this, most models have still little to no connection to empirical data. Furthermore, many problems about data are still unresolved – it is still not clear if toy models could be used with empirical data, which models may be more robust to psychometric distortions or even how to deal with change of scales. Because of this, more research effort needs to be dedicated to the connection between these models and empirical data. This could even help with the previous problem, as some measurables (e.g. a specific measurement scale) could become common ground between different fields. For example, it might be very productive if multiple models of the same situation and data were made and compared.

Encouraging more complicated modelling. As mentioned, most of the research in social simulations is focused on relatively simple models that aim to understand the possible effects of one kind of mechanism in a generic manner. Because of that, a lot of effort is dedicated to the development of new kinds of model or to the exploration of alternatives to established models. Meaning that, contrary to other fields such as physics, researchers usually do not keep developing and testing existing models but keep developing new ones. While this allows for a faster exploration of new possible effects, the sheer proliferation of models can make the field more inaccessible to outsiders. Furthermore, the lack of empirical validation makes it hard to decide the most appropriate model for a specific context of scenario. Models which can provide insights on how to "get people to agree" in realistic situations will probably be complicated and integrate many different types of interaction and socio-cognitive processes. Therefore, we need to develop an environment that allows and even encourages models to incorporate whatever mechanisms or details are required to adequately represent the observed cases in empirical terms. These can usefully be compared to more abstract, single mechanism models, but not be restricted to them.

6 Conclusion

The social simulation community is in a unique position to contribute towards helping humanity survive future crises by better understanding how to help people agree, and so contribute to the development of collective initiatives to avoid and mitigate the crises that we face. However, this will require model families that bridge those in current subfields, including multi-mechanism models to explore conditions of application and have a much stronger relationship with empirical data from observed interactions. This will, itself, require a collective effort, but we do not have time to reach an agreement on such a project – we have to get going on it immediately!

Acknowledgements

BE and ECB were supported as part of the ESRC-funded, UK part of the "ToRealSim" project, 2019-2023, grant number ES/S015159/1. DC is grateful for support by the project "CoCi: Co-Evolving City Life", which received funding from the European Research Council (ERC) under the EU's Horizon 2020 programme, grant 83316.

References

1. Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. Journal of conflict resolution, 41(2), 203-226.
2. Boxell, L., Gentzkow, M., & Shapiro, J. M. (2022). Cross-country trends in affective polarization. Review of Economics and Statistics, 1-60. DOI:10.1162/rest_a_01160/109262
3. Carpentras, D., & Quayle, M. (2023). The psychometric house-of-mirrors: the effect of measurement distortions on agent-based models' predictions. International Journal of Social Research Methodology, 26(2), 215-231.
4. Carpentras, D. & al. (2022). Deriving An Opinion Dynamics Model From Experimental Data. Journal of Artificial Societies and Social Simulation, 25(4), 4. https://jasss.org/25/4/4.html
5. Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. Reviews of Modern Physics, 81(2), 591.
6. Centola, D. (2022). The network science of collective intelligence. Trends in Cognitive Sciences. 26(11), 923-941.
7. Conte, R., & Castelfranchi, C. (2006). Understanding the functions of norms in social groups through simulation. In Artificial Societies (pp. 225-238). Routledge.
8. Conte, R., Andrighetto, G., & Campennl, M. (Eds.). (2014). Minding norms: Mechanisms and dynamics of social order in agent societies. Oxford University Press.
9. Deffuant, G. & al. (2000). Mixing beliefs among interacting agents. Advances in Complex Systems, 3(01n04), 87–98.
10. DeGroot, M. H. (1974) Reaching a consensus, Journal of the American Statistical Association, 69(345), 118-121.
11. Dignum, F., Edmonds, B. & Carpentras, D. (2022) Socio-Cognitive Systems – A Position Statement. Review of Artificial Societies and Social Simulation, 2nd Apr 2022. https://rofasss.org/2022/04/02/scs/
12. Edmonds, B. (2005). Assessing the safety of (numerical) representation in social simulation. 3rd ESSA conference, Koblenz, Germany. http://cfpm.org/cpmrep153.html

13. Edmonds. B. & Hales, D. (2004) When and Why Does Haggling Occur? Some suggestions from a qualitative but computational simulation of negotiation. Journal of Artificial Societies and Social Simulation, 7(2), 9. http://jasss.org/7/2/9.html
14. Flache, A. & al. (2017). Models of social influence: Towards the next frontiers. Journal of Artificial Societies and Social Simulation, 20(4), 2. http://jasss.org/20/4/2.html
15. Goldberg, A., & Stein, S. K. (2018). Beyond social contagion: Associative diffusion and the emergence of cultural variation. American Sociological Review, 83(5), 897-932.
16. Gromet, DM, H Kunreuther, H & Larrick, RP (2013) Political ideology affects energy-efficiency attitudes and choices. Proceedings of the National Academy of Sciences USA 110, 9314–9319.
17. Heidari, S. (2022) Agents with Social Norms and Values: A framework for agent based social simulations with social norms and personal values. Thesis, Utrecht NL. https://dspace.library.uu.nl/handle/1874/422431
18. Hołyst, J. A., Kacperski, K., & Schweitzer, F. (2001). Social impact models of opinion dynamics. Annual Reviews Of Computational Physics, IX, 253-273.
19. Ito, T. & al. (Eds.). (2009). Advances in agent-based complex automated negotiations (Vol. 233). Springer.
20. Jager, W. & Amblard, F. (2005). Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. Computational and Mathematical Organization Theory, 10, 295-303.
21. Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. Proceedings of the national academy of sciences, 108(22), 9020-9025.
22. Malone, T. W., & Bernstein, M. S. (Eds.). (2022). Handbook of collective intelligence. MIT press.
23. Martins, A. C. (2008). Continuous opinions and discrete actions in opinion dynamics problems. International Journal of Modern Physics C, 19(04), 617-624.
24. Oakes, P. J., Haslam, S.A. and Turner, J.C. (1994). Stereotyping and social reality. Blackwell Publishing.
25. Santos, F.C. & al. (2012). The role of diversity in the evolution of cooperation. Journal of Theoretical Biology 299, 88–96.
26. Scholz, G., & al. (2023) Social Agents? A Systematic Review of Social Identity Formalizations. Journal of Artificial Societies and Social Simulation, 26(2), 6. http://jasss.org/26/2/6.html
27. Schut, M. C. (2010). On model design for simulation of collective intelligence. Information Sciences, 180(1), 132-155.
28. Tajfel, H. (ed.) (1978) Differentiation Between Social Groups: Studies in the Social Psychology of Inter-group Relations. London: Academic Press.
29. Turner, J.C. & al. (1987) Rediscovering the Social Group: A Self-Categorization Theory. Blackwell
30. Ullmann-Margalit, E. (1977). The Emergence of Norms. Oxford University Press.
31. Van Boven, L., & Thompson, L. (2003). A look into the mind of the negotiator: Mental models in negotiation. Group Processes & Intergroup Relations, 6(4), 387-404.
32. Xenitidou, M. & Edmonds, B. (Eds.) (2014) The Complexity of Social Norms. Springer.