Mathematical Social Sciences 80 (2016) 97–107

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

Opinion dynamics and wisdom under out-group discrimination

Steffen Eger

Ubiquitous Knowledge Processing Lab, Computer Science Department, Technische Universität Darmstadt, Germany

HIGHLIGHTS

- A DeGroot-like opinion dynamics model with in- and out-group structure is proposed.
- Agents want to coordinate with their in-group & anti-coordinate with their out-group.
- Model predictions include persistent disagreement in connected societies.
- Necessary and sufficient conditions for bi-polarization are determined.
- We also determine opinion leadership and discuss (crowd) wisdom.

ARTICLE INFO

Article history:
Received 26 September 2015
Received in revised form
16 February 2016
Accepted 19 February 2016
Available online 2 March 2016

ABSTRACT

We study a DeGroot-like opinion dynamics model in which agents may oppose other agents. As an underlying motivation, in our setup, agents want to adjust their opinions to match those of the agents of their 'in-group' and, in addition, they want to adjust their opinions to match the 'inverse' of those of the agents of their 'out-group'. Our paradigm can account for persistent disagreement in connected societies as well as bi- and multi-polarization. Outcomes depend upon network structure and the choice of deviation function modeling the mode of opposition between agents. For a particular choice of deviation function, which we call soft opposition, we derive necessary and sufficient conditions for long-run polarization. We also consider social influence (who are the opinion leaders in the network?) as well as the question of wisdom in our naïve learning paradigm, finding that wisdom is difficult to attain when there exist sufficiently strong negative relations between agents.[1]

1. Introduction

On many economic, political, social, and religious agendas, disagreement among individuals is pervasive. For example, the following are or have been highly debated: whether abortion, gay marriage, or death penalty should be legalized or not; whether Iraq had weapons of mass destructions; the scientific standing of evolution; whether taxes/social subsidies/unemployment benefits/(lower bounds on) wages should be increased or decreased; the right course of government in general; the effectiveness of alternative (or 'standard') medicine such as homeopathy.[2] In fact, in certain contexts such as the political arena, disagreement is 'built into' and essential part of the system of opinion exchange (Jones, 1995; Cohen, 2003). Yet, contradicting this factual basis, it has been observed that the phenomenon of disagreement is not among the predictions of renown and widely used theoretical models of opinion dynamics in the social and economic context, whether they are based on fully rational, *Bayesian*, agents or boundedly rational or *non-Bayesian* actors.[3] Namely, in these models, a standard prediction is that agents tend toward a *consensus opinion*, that is, all agents eventually hold the same opinion (or belief)[4] about any specific issue. Typically, this applies to both Bayesian frameworks – which is the reason why Acemoglu and Ozdaglar (2011) call them "[no] natural framework[s] for understanding persistent disagreement" (p. 6) – and non-Bayesian setups such as the famous DeGroot model of opinion dynamics DeGroot (1974), where consensus

E-mail address: eger.steffen@gmail.com.

[1] Earlier and more verbose working paper versions of this article can be found at http://arxiv.org/pdf/1306.3134 and the author's personal website.

[2] Our examples are, i.a., taken from Abramowitz and Saunders (2005), Acemoglu and Ozdaglar (2011) and Golub and Jackson (2012).

[3] See, e.g., the discussions Acemoglu and Ozdaglar (2011) and Acemoglu et al. (2013). See also Abelson (1964).

[4] Typically, in the literature, the term *belief* is used when there exists a *truth* for an agenda, and the term *opinion* is used when truth is not explicitly modeled. Like our related work, we more generally subsume under the term opinions also beliefs, judgements, estimations, or even norms and values, depending on the application scenario.

http://dx.doi.org/10.1016/j.mathsocsci.2016.02.005
0165-4896/

obtains as long as the social network wherein agents communicate with each other is strongly connected (and aperiodic).

Concerning the non-Bayesian DeGroot model, as we consider in this work, a few amendments have more recently been suggested which are capable of producing disagreement among agents. In one strand of literature, models including a *homophily* mechanism, whereby agents limit their communication to individuals whose opinions are not too different from their own, can reproduce patterns of opinion diversity and disagreement (Deffuant et al., 2000; Hegselmann and Krause, 2002). In another strand, Daron Acemoglu and colleagues (cf. Acemoglu and Ozdaglar, 2011) introduce two types of agents, *regular* and *stubborn*, whereby the latter never update their opinions but 'stubbornly' retain their old beliefs, which may be considered an autarky condition. Multiple stubborn agents with distinct opinions on a certain agenda may then draw society toward distinct opinion clusters. Such stubborn agents, it is argued, may appear in the form of opinion leaders, (propaganda) media, or political parties that wish to influence others without receiving any feedback from them. As a solution to the disagreement problem, however, both of these model types rely on a problematic 'disconnectedness condition', insofar as disagreement only obtains when there is no (uni- or bilateral) information flow between certain subclasses of agents.

In this work, we investigate an alternative explanation of disagreement, which can also explain disagreement in connected societies. We consider a non-Bayesian DeGroot-like opinion dynamics model in which agents are related to each other via *two types of links*. One link type represents the *degree* or *intensity* of relationship between agents and is given by nonnegative real numbers. The other link type represents whether agents *follow* or *oppose* (*deviate from*) each other, that is, it represents the *kind of relationship* between agents. We assume that *group identity* causes agents to follow their *in-group* members and to deviate from their *out-group* members. In-group favoritism and out-group discrimination are important and well-established notions in social psychology (see, for instance, Tajfel et al., 1971; Brewer, 1979 and Castano et al., 2002). They have also more recently been included in economists' models (e.g., in an experimental context, Charness et al., 2007, Ben-Ner et al., 2009, Chen and Li, 2009, Hargreaves Heap and Zizzo, 2009, Fehrler and Kosfeld, 2013 and Tsutsui and Zizzo, 2014). Experimentally, it has been shown that even minimal group identities, induced by a random labeling of groups, may lead to intergroup discrimination. When group membership is more salient, Charness et al. (2007) show that there is much more cooperation between in-group members than between out-group members in a prisoner's dilemma game, and Fehrler and Kosfeld (2013) show that individuals associating with particular NGOs (non-governmental organizations) strongly discriminate against out-group members (those that do not associate with an NGO) in a trust game.[5] Analogously, we assume that agents want to coordinate with their in-group members (have negative utility from holding different opinions than in-group members) and want to anti-coordinate with their out-group members (have negative utility from not deviating from the opinions that their out-group members hold).[6] A special case of our model is when an agent opposes everyone but himself, i.e., his in-group is himself and his out-group is all 'the rest'. In some works, such agents have been referred to as *rebels* or *anti-conformists* in contrast to *conformists* (Jackson, 2009; Cao et al., 2011; Javarone, 2014; Jarman et al., 2015).

Our model closely follows the literature on learning through communication in a given social network (cf. DeGroot, 1974, DeMarzo et al., 2003, Golub and Jackson, 2010, Acemoglu et al., 2010 and Buechel et al., 2015). There, the standard assumption is that agents learn from others in a naïve manner, not properly accounting for the repetition of opinion signals, which DeMarzo et al. (2003) call 'persuasion bias'. A now classical argument is that if all agents' initial beliefs/opinions were independent and unbiased estimates of the true value (of a discussion topic), then taking a weighted average of the agents' beliefs in one's social network (where the weights are proportional to the inverses of the beliefs' variances) is an optimal aggregation strategy. Then, continuing to average – in order to incorporate more remote information, e.g., from friends of friends – *in the same manner* is a boundedly rational heuristic that treats the evolving information signals as novel, not accounting for their cross-contamination. Such a heuristic aggregation of opinion signals appears quite plausible given the processing costs involved in exact inference in this setup (cf. Golub and Jackson, 2010). Also, recent experimental evaluations find that the naïve DeGroot model is a much better approximation of information aggregation in network interactions than 'fully rational' Bayesian approaches and that individuals are indeed affected by persuasion bias (Corazzini et al., 2012). In our model, we posit that agents are subject to the same biases involving processing of cross-contaminated information and are, *in addition*, susceptible to *in-group bias*, attempting to coordinate with in-group members and to anti-coordinate with out-group members.[7]

This work is structured as follows. Section 2 presents the model and gives two introductory examples. Sections 3 and 4 present our main results, on persistent disagreement (Theorem 3.1) and bi- and multi-polarization (Proposition 3.2). For a special case of our model, we derive necessary and sufficient conditions for long-run polarization (Theorem 4.1) as well as opinion leadership (Theorem 4.2) as further main results. In Section 5, we summarize and conclude with a discussion on wisdom. To make this work more or less self-contained, we provide concepts, e.g., from graph and matrix theory in the Appendix, to which we also relegate all our proofs.

2. Model

Let S be a finite set ('discrete model') or a subset of the real numbers ('continuous model'), which we refer to as *opinion spectrum*.[8] Let $n \geq 1$ and let $[n] = \{1, \ldots, n\}$ be a set of n agents. Consider the *normal form game* $([n], S_1 \times \cdots \times S_n, U)$, where

- $[n]$ is the set of *players*,

[5] In a 'field' setting, the in-group/out-group distinction may prominently be seen as arising, e.g., in a (main stream) culture/counterculture (e.g., hippies, punks, etc.) dichotomy (Yinger, 1977) or in classical party divisions (e.g., Republicans vs. Democrats) in the field of politics, etc.

[6] Out-group discrimination (opposition) is also closely related to what has been termed *rejection* of beliefs, actions, and values of dissimilar/disliked others. According to this concept, agents change their normative systems to become more dissimilar to interaction partners they dislike (cf. Abelson, 1964; Kitts, 2006 and Tsuji, 2002; cf. also Groeber et al., 2013) insofar as disliked others may serve as 'negative referents' who inspire contrary behavior. While in controlled experiments Takács et al. (2014) do not find strong evidence for the tenet that individuals disassociate from the opinions of a disliked source, their study explicitly excludes a group identity structure. Moreover, as the authors argue, their laboratory experiment may have "suppressed the emotional processes that in field settings induce disliking and rejection of others' opinions".

[7] If one wanted to construct an argument that closely follows that of DeMarzo et al. (2003), one might posit that, in our model, agents 'correct' their out-group members' opinion signals – possibly because of distrust – before averaging.

[8] For the continuous DeGroot model as we discuss, S is typically modeled as a convex subset of the real numbers, that is, $\sum_j \alpha_j x_j \in S$ for all finite numbers of elements $x_j \in S$ and all weights $\alpha_j \in [0, 1]$ such that $\sum_j \alpha_j = 1$. For convenience, we think of S as the whole of $\mathbb{R}$ or of some (closed and bounded) interval $[\alpha, \beta]$ for $\alpha \leq \beta$.