The diversity puzzle

Mäs, Michael

I. Overview of the book

I.1. Social-influence processes in human groups

Example 1: Reaching consensus in the jungle.

Your plane crashed in the jungle. When the plane hit the ground, all crew members were killed but you and three other passengers survived. The group of survivors is demographically mixed. There are two males and two females. Two of you are black and two are white. After having waited for days on end in the jungle, you still have not been rescued. The four of you discuss what to do and agree that the group should try to find a way out of the jungle. The problem is that no one has a clue in which direction to go. One of the other passengers asks you: "Listen, as a sociologist you should be an expert on this. Is it wise to discuss the problem or should one of us decide for the group? To maximize our chance of survival, we have to remain together as one group. Finding a consensus would be optimal because all of us would support the decision. However, when we discuss the problem as a group, people might disagree and someone might pick a fight. This is the worst thing that can happen in our situation. Alternatively, one of us may assume the leadership and decide for the group. This would be better than a conflict but certainly worse than the consensus." What would your advice be? How likely is it that in a discussion all the survivors can be convinced that one given direction is the most promising? How would the group's demographic diversity affect the discussion? Will it make conflicts more or less likely?

Example 2: The Oscars.

Every year, the Academy of Motion Picture Arts and Sciences nominates five actors for the Oscar for the Best Actor for that year. On the night when the awards are presented, everybody who works in the movie business is in Hollywood. When the winner is announced, all these people will have an opinion about this decision and it is likely that opinions will vary considerably.

> Some will believe that the winner very much deserved the award. Others will feel that the winner's work is overrated and another nominee deserved the award more. Later, at the famous Oscar parties, people will talk about the Academy's decision. They will discuss why they think that the winner deserves the Oscar or why not. As a result of these discussions, will the opinions of the movie people converge? In other words, will there be consensus on whether the winner deserves the prize or will opinions remain different?

Example 3: How cool is rap music?

> At the beginning of high school, a student becomes involved in rap music. Her class mates are interested in music but do not have a clear-cut musical preference yet. How will the musical preferences of the other students develop in the subsequent months? Will the rap fan infect the others? Under what conditions will the others develop different musical preferences?

These examples have two central aspects in common. First, each member of the various groups holds an *opinion* on a certain issue. In other words, each of them has evaluated a certain object (a certain direction to go, the decision of the Academy, rap music). Opinions vary between two extremes and can be measured on a metric scale. For example, the opinion of each plane-crash survivor can be quantified using the degrees on the compass rose. Each survivor prefers a specific direction. The more degrees that the preferred direction of another survivor differs from this direction, the more the two actors disagree. Second, the members of each group interact and may *influence* each others' opinions. In particular, they may persuade each other and adjust their opinions in such a way that they become more similar.

Research shows that social influence plays an important role in social interaction. It is argued, for instance, that interaction partners exchange arguments and persuade each other that certain opinions are more adequate (Myers 1982; Wood 2000). Other theorists argue that interaction partners may exert social pressure to conform with each other (Festinger, Schachter and Back 1950; Homans 1951). Furthermore, cognition theories (Festinger 1957; Heider 1967) imply that we want to be similar to people we like to interact with. To achieve this, we might try to convince our friends to adopt opinions and behavior similar to ours, or, the other way around, we might change our opinions and attributes to conform to those of our friends. Furthermore, social influence may also result from imitation (Akers et al.

1979). It has been argued, for instance, that in situations of high uncertainty it can be rational for individuals to imitate the behavior and opinions of others (Bikhchandani, Hirshleifer and Welch 1992).

Empirical studies also demonstrate the importance of social influence. For instance, psychological experiments consistently show that subjects adjust their opinions after having been informed about the opinions of another person (see Wood 2000 for an overview). In particular, when subjects know that they share some attribute with that person, they tend to decrease their opinion distance (e.g. Berscheid 1966; Sampson and Insko 1964; van Knippenberg and Wilke 1988). In addition, research has been conducted on the outcomes of discussions in social groups. In these experiments, typically groups of four subjects meet and discuss some issue (Johnson and Johnson 1982). It has been found that, as a result of the discussion, these participants adjust their opinions (Myers 1982; Wood 2000).

This book is concerned with explaining the opinion dynamics that social influence causes in social groups. Based on these explanations, we seek to predict the distribution of opinions which result from social influence processes. Under what conditions will the members of a group find a consensus? Under what conditions will initial opinion differences persist? Is it possible that opinion differences between subgroups will increase over time? Is opinion consensus always stable, or can homogeneous groups split up into subgroups with opposing opinions? Our goal is to develop theories which can be applied to all social groups, from small groups such as work teams and school classes to large scale groups such as social networks on the Internet.

Our focus in this book is on finding an *explanation* for social influence dynamics. We developed theories to describe opinion dynamics that result from social influence and to generate predictions about the outcomes of influence processes. It has been argued that theories of social-influence may also help us understand under what conditions groups find the best or the correct answer to a given problem (see for instance Kennedy 1998). It has also been proposed that management teams arrive at better decisions than individual managers (Johnson and Johnson 1982). Following a similar line of reasoning, it has been claimed that there is something called the *wisdom of the crowds* (Surowiecki 2005) in the sense that aggregating the information of crowd members results in better decisions than the decisions from single crowd members. That particular topic, however, is beyond the scope of this book. Before we study under what conditions groups make good decisions, we need to have a reliable theory of opinion formation in groups. We will show in the following

section that existing models have several shortcomings and that therefore our focus will be on elaborating on the existing theories.

I.2. Existing theories of social-influence dynamics and their shortcomings

In this section we will review the existing theories which have been developed to understand social-influence dynamics. We will summarize the answers that these theories give to the questions we have just formulated and point to their shortcomings. Finally, we will formulate two research questions which the existing models fail to answer.

I.2.1. Classical social-influence models

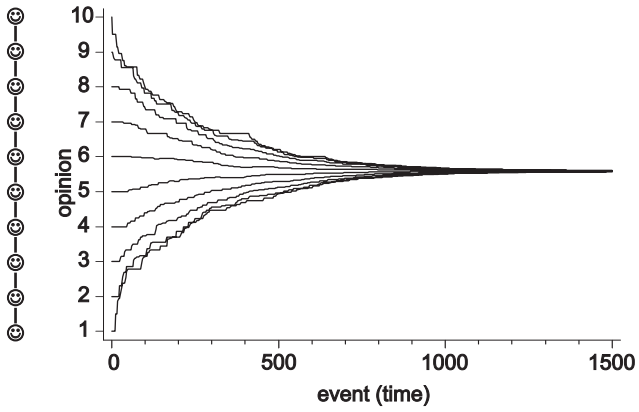
Already in the 1950s, researchers have developed formal theories of social-influence processes in groups (Abelson 1964; Berger 1981; DeGroot 1974; French 1956; Friedkin and Johnsen 1990; Harary 1959; Lehrer 1975). These models express the opinion of each group member as a real number which can vary between two extremes. As an example, this value might reflect the opinion on rap music of one of the students from Example 3. Small values would indicate a negative opinion and high values a positive opinion.

In contrast to the classical literature on social-influence dynamics, several models have been developed which assume that actors influence each other on *nominal* variables (Axelrod 1997; Carley 1991; Galam 2005; Liggett 1985; Mark 1998; Mark 2003; Nowak, Szamrej and Latané 1990; Sznajd-Weron and Sznajd 2000). These nominal variables represent for instance whether the actors adopt a piece of information (Carley 1991) or the political party the actors will vote for (Liggett 1985). However, many of the attributes which are open to social influence are better described using metric scales. For instance, very few people agree or disagree exactly with the program of a political party. Instead, people usually agree to a certain degree. We will show in this section that existing social-influence models are problematic when they are applied to such continuous dimensions. Our aim is to overcome these problems. We will therefore stick to the assumption of the classical social-influence models and focus on continuous opinions. We should also point out that nominal variables can not serve as an approximation of metric attributes. This is because predictions of social-influence models critically depend on whether actors influence each other in terms of nominal or continuous dimensions (Flache and Macy 2006a).

When incorporating the social-influence mechanism, modelers assume that the actors adjust their opinions in such a way as to become more similar to the other group members.

In particular, they assume that actors adopt an opinion equal to the weighted average of the others' opinions. As a result, models take into account the fact that not all group members are equally influential for an actor. The impact that an actor *j* has on actor *i*'s opinion is expressed by a weight $w_{i,i}$. A weight of zero represents the fact that *j* has no influence on *i*'s opinion. Values above zero indicate how strong the impact of *j* is on *i*. Classical models assume that weights do not change over time.

Figure I.1: Opinion trajectories of 10 actors in the classical social-influence model



To illustrate the opinion dynamics which follow from these assumptions, we show in Figure I.1 a typical social-influence process in a group of 10 actors, named 1 thru 10. At the outset of the influence dynamics, each actor holds an opinion the same as his name. Furthermore, the group has a very simple influence network: a line. That is, Actor 1 is only influenced by Actor 2 ($w_{1,2}$=1). Actor 2 is only influenced by Actors 1 and 3 ($w_{2,1}$= $w_{2,3}$=1). Actor 3 is only influenced by Actors 2 and 4 ($w_{3,2}$= $w_{3,4}$=1) and so on. The remaining weights (e.g. $w_{1,3}$) are zero. For the sake of illustration, we have modeled the group's social-influence process as a sequence of events. At each event, we picked one of the ten actors and adjusted his opinion. In line with classical social-influence models, we assigned a new opinion that was similar to the average opinion of the actor and his network partners.

Figure I.1 demonstrates that the opinions of the actors converge and become basically identical in the long run. At the beginning of the process, Actors 2 thru 9 do not change their opinions because the influences from their two contacts balance each other out. However, Actors 1 and 10 are influenced by only one group member. In both cases, the opinion of the contact is less extreme than the actor's own opinion. As a consequence, Actors 1 and 10 develop more moderate opinions. This, in turn, affects the opinions of Actors 2 and 9. Because their extreme contact (1 and 10, respectively) has developed a

more moderate opinion, the opinions of 2 and 9 also shift towards a less extreme view. These changes, in turn, trigger opinion adjustments for Actors 3 and 8 and so on.

It has been shown that the result of the example from Figure I.1 can be generalized. More precisely, researchers have proved analytically that classical social-influence models *always* predict convergence of opinions towards global uniformity unless there is a subset of actors that is completely cut-off from outside influences (Abelson 1964; Berger 1981; DeGroot 1974; Harary 1959; Lehrer 1975; Wagner 1982). Thus, irrespective of the structure of the influence network and the initial opinion distribution, consensus appears to be unavoidable. According to the classical models, consensus formation can only be stopped by social separation.

This result is puzzling because it contrasts with the diversity of opinion that we observe in real life. For instance, research on group dynamics in work teams demonstrates that team members often fail to decrease their differences of opinion within the team (van Knippenberg and Schippers 2007). Members of work teams interact frequently and are therefore easily able to influence each other. Yet, differences of opinion do not always decrease; sometimes they even intensify (Early and Mosakowski 2000).

In a similar way, long-term studies on the social attitudes of Americans show that there is significant opinion variation on relevant issues such as feelings towards blacks, views on gender equality and the evaluation of abortion (DiMaggio, Evans and Bryson 1996). Studies did find decreasing variance for many opinion dimensions (e.g. opinions on matters related to crime and justice). However, studies also found that the variance of several opinion dimensions has *not* changed significantly from the beginning of the 1970s to the beginning of the 1990s (DiMaggio, Evans and Bryson 1996; Mouw and Sobel 2001). A study which included data up until the year 2000 found that variance has actually increased for dimensions such as sexual morality and attitudes towards poor people (Evans 2003). There is also research showing that the American diversity in terms of lifestyles and consumption tastes has increased (for an overview: Fischer and Mattson 2009).

Social separation does not provide a plausible explanation for these findings. Modern societies have developed very efficient transportation and communication technologies which make it very easy to be in contact even with geographically distinct people (Greig 2002). Furthermore, modern "worlds" are surprisingly small (Milgram 1967). It has been demonstrated that in modern societies most pairs of individuals seem to be connected by a very short path through the network (Barabasi 2003; Travers and Milgram 1969; Watts

2003; Watts and Strogatz 1998). This suggests that in modern societies individuals are exposed to influence from various directions. Yet, diversity of opinion remains high.

An alternative explanation for opinion diversity may be that some opinions are related to material interests. For instance, rich people typically hold negative opinions on issues like inheritance tax increases. Such interests may interfere with social-influence processes and bring opinion convergence to a stop when actors refuse to adjust their opinions because this would contradict their material interests. In line with this argument, empirical research in the US has found opinion differences between respondents with a college degree and respondents with no more than a high-school education (DiMaggio, Evans and Bryson 1996). However, this research also showed that opinion differences between the two groups decreased between the 1970s and the early 1990s. Furthermore, these differences can not explain opinion diversity on issues that are unrelated to material interests such as sexual morality (Evans 2003) and the attitude towards abortion (DiMaggio, Evans and Bryson 1996).

A related point has been made by Friedkin and Johnson (1997) who assumed that actors seek to adopt opinions which mirror their views at the outset of the influence process. Their social-influence model is able to generate opinion dynamics with decreasing opinion variance at the beginning of the process but stable opinion diversity in the long run. Dynamics reach equilibrium when social pressures to conform to others' opinions and the striving towards the initial opinion cancel each other out such that opinions remain stable. We do not question this assumption. However, we would like to point out that this model can not generate the *increasing* opinion diversity which empirical studies have found on issues such as sexual morality and attitudes towards poor people (Evans 2003). Moreover, the Friedkin-Johnson model can generate social-influence dynamics that result in opinion diversity only in those settings that are already at the outset of the influence process characterized by high opinion diversity. This model does not, however, offer an explanation for the emergence of opinion diversity. In other words, the model predictions rely on the existence of opinion distributions which the model cannot generate endogenously. In this book, we will develop models which do not have these shortcomings.

In summary, classical social-influence models predict that opinions will converge over time. However, there is a conflict between this prediction and the opinion dynamics which we observe empirically. Puzzled by this, Abelson wondered already in 1964 "what on earth one must assume in order to generate the bimodal outcome of community cleavage

studies?"(153). With this, Abelson identified a puzzling and tenacious research problem. In this book, we will follow Abelson's lead and seek to explain why there is persistent opinion diversity despite social influence.

I.2.2. One potential solution: Selection of similar interaction partners

One potential solution to Abelson's puzzle has been proposed in the literature on cultural and group dynamics (Axelrod 1997; Carley 1991), in the literature on decision-making in evolving networks (Stokman and Zeggelink 1996a; Stokman and Zeggelink 1996b) and in the social-influence literature (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002). The assumption found here is that actors tend to interact with those others who hold similar opinions and avoid influence from dissimilar others.

This notion is supported by two prominent theories (Byrne 1971). First, the reinforcement approach (Byrne 1961) argues that humans do *not* have a general predisposition for interacting with similar others. However, whenever we interact with similar others we feel rewarded. When an interaction partner "offers us validation by indicating that his percepts and concepts are congruent with ours, it constitutes a rewarding interaction and, hence, one element in forming a positive relationship" (Byrne 1961: 713). At the same time, interaction with dissimilar others constitutes a negative stimulus because we learn that our opinions may be wrong. Furthermore, such interactions may end up in punishing discussions and conflicts. In other words, during interaction with similar and dissimilar others we learn that meeting with similar others just feels better and, hence, we start to prefer such interaction partners.

A second explanation is based on cognitive theories. These theories argue that humans strive for balanced cognitions (Festinger 1957; Heider 1967). This is achieved when we have positive emotions towards persons who are similar to us, and when we have negative emotions towards persons who are dissimilar. For instance, the two cognitions "I *dis*like Tom" and "I agree with Tom" are unbalanced. According to these cognitive theories, unbalanced cognitive constellations are unpleasant for us and we try to avoid them. In the example, this can be achieved by changing our relationship to Tom from disliking to liking. In short, when we learn that an interaction partner holds similar opinions, it just feels good to like him.

Empirical research also supports the assumption that individuals tend to interact with similar others. Social psychological experiments along the Similarity-Attraction-Paradigm (for a comprehensive summary see Byrne 1971) support the fact that humans evaluate

similar others more positively than dissimilar others. In these experiments, subjects are introduced to a stranger along with information on the stranger's opinions. Experiments show consistently that subjects rate strangers more positively the more they agree with the opinion of the stranger. In addition, research on the determinants of friendships shows that humans tend to nominate similar group members as their friends. To illustrate, it has been found that students prefer friends with similar smoking behavior (Mercken et al. forthcoming). Similar effects have also been found for delinquency (Burk, Steglich and Snijders 2007) and alcohol consumption (Pearson, Steglich and Snijders 2006).

Researchers have included the notion that individuals tend to select similar others as interaction partners in social-influence models. In particular, it is assumed that actors refuse to be influenced by actors who are too dissimilar. The maximum dissimilarity that actors consider acceptable is called the confidence level. Accordingly, such models are called *models with bounded confidence* (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002; Lorenz 2007). To incorporate this notion into the classical social-influence framework, it is implemented that the influence weights are a function of the opinion similarity between the actors. In particular, weights are assumed to be zero as the dissimilarity between two actors exceeds a certain threshold; otherwise, they are one[1] (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002).

The interplay of the selection of similar others and social influence creates a feedback process which can lead to the development of distinct clusters of similar actors (Axelrod 1997; Carley 1991; Hegselmann and Krause 2002). In this process, similar actors interact

[1] To be more precise, models of decision making in evolving networks (Stokman and Zeggelink 1996a; Stokman and Zeggelink 1996b) are based on a slightly different implementation of the bounded-confidence assumption. In these models, actors influence each other only if they have established a mutual network relationship. Furthermore, the probability that an actor will attempt to create such a relationship and that this attempt will be reciprocated depends on the opinion similarity between the two actors. It is thus *un*likely that dissimilar actors will establish a mutual network relationship. Furthermore, these models assume that, if an actor *i* attempts to create a relationship but the potential network partner *j* rejects this offer, then the probability that *i* will repeat the attempt to create a network relationship to *j* at a later time decreases. Since rejecting network ties is more likely the more the opinions of *i* and *j* differ, the probability that dissimilar agents will create a relationship very likely decreases quickly and eventually adopts zero. This rules out any future influence between these actors.
Note that some models of decision-making in evolving networks (Stokman and Zeggelink 1996b) combine the bounded-confidence assumption with an additional mechanism that may prevent opinion convergence even in a connected network. Contrary to the classical social-influence models and the models which we will develop in the following chapters, some models of decision making in evolving networks assume that opinions can adopt only one of two values, zero or one. In addition, it is assumed that actors adopt an opinion which is similar to the median of all network partners' opinions (Stokman and Zeggelink 1996b: page 401 and footnote 15). As a consequence, an actor *i* may have a relationship to an actor *j* who holds an opinion value of one. However, if there is a sufficient number of other network partners who hold the opposite opinion (zero), then *i* will not be influenced by *j* and will adopt an opinion of zero.

and thereby become even more similar. This, in turn, leads to further interaction which again increases similarity. In this process, distinct clusters of similar actors can emerge. Actors at the center of a cluster pull actors from the border of their cluster closer to themselves. In doing so, they pull them away from other clusters. Subsequently, the members of a cluster become more and more similar and collectively ignore actors outside their cluster.

To illustrate how the interplay of selection and influence triggers cluster formation, Figure I.2 shows a scenario which we have replicated from Hegselmann and Krause's computer simulation study (2002). In this simulation, we modeled the opinion dynamics of 100 actors that hold random opinions at the outset. We imposed that the influence weights adopt the value one for pairs of actors with a dissimilarity smaller than 20% of the width of the opinion scale and zero otherwise.

Figure I.2: Opinion trajectories of 100 actors generated by the interplay of selection and social-influence (BC-model of Hegselmann and Krause (2002))

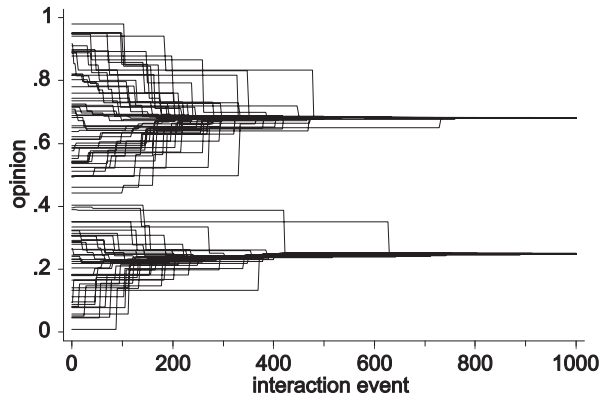


Figure I.2 shows that initially opinion is uniformly distributed. In contrast to the classical social-influence models (see Figure I.1), however, two clusters form. The dissimilarity between the clusters exceeds the imposed dissimilarity threshold of 20%. As a consequence, actors that belong to distinct clusters ignore each other. Dynamics settle when the clusters are internally perfectly homogeneous.

As Figure I.2 illustrates, the feedback process generated by the interplay of selection and influence can lead to opinion clustering and can stop the inevitable march towards consensus that classical social-influence models predict. However, we argue that this approach relies heavily on two problematic assumptions. First, the bounded confidence assumption proposes that actors refuse to interact with others who are too dissimilar. We

do not question this assumption in general but it may be too strict to assume that dissimilar actors never ever interact. In Chapter V (see Figure V.1, Panel B) we show in more detail that including even an extremely small probability of interaction between actors who are too dissimilar to interact otherwise suffices to reactivate the march towards consensus of the classical models. As an illustration, let us take a look at the opinion distribution at the end of the process shown in Figure I.2. There are two distinct clusters which are sufficiently dissimilar to rule out further influence. But, if an actor happens to be influenced by the opposing cluster, this actor will move slightly nearer to this cluster. Subsequently, the actor will influence the other members of his own cluster and thereby pull them nearer to the opposing group. If such events occur more often, the clusters will gradually move nearer to each other and eventually be similar enough to exert influence on each other and merge. In real life, such events may seldom occur. However, if they can not be ruled out completely, then bounded-confidence models predict opinion convergence, too.

A second shortcoming of the bounded-confidence model is that it can generate clustering only when there is sufficient opinion variation already at the beginning of the influence process. In Figure I.2, for instance, the dynamics started with a uniform opinion distribution. Such populations are comprised of extreme actors who will pull others towards them. This is crucial because without a sufficient number of extremists, emerging clusters will be too similar and therefore merge in the long run. This problem is rooted in the nature of social influence. It implies that interaction partners become more similar. It follows that diversity can only decrease over time. Thus, there is no endogenous mechanism that can create diversity. The bounded-confidence explanation of opinion diversity is problematic because it hinges on the existence of sufficient initial diversity, while at the same time the model fails to generate that diversity.

In sum, researchers have included the assumption that actors will tend to select similar interaction partners and refuse to interact with group members who are too different in their opinions. We argue that this is in fact a promising strategy for two reasons. First, this assumption is supported by empirical research. Second, it has been shown that the resulting models are capable of explaining opinion diversity. However, the new models fail at the same time to explain opinion diversity when only very small perturbations from the bounded-confidence assumption are included. This, and the fact that the models fail to create diversity, demonstrates that the bounded-confidence models still do not yet offer a satisfactory explanation for opinion diversity.

I.2.3. Research Questions

This brief review of existing social-influence models has identified a conflict between the models' predictions and empirical evidence. One the one hand, existing social-influence models imply that social influence triggers convergence cascades which lead to decreasing opinion diversity and eventually end up with perfect uniformity. Empirical evidence, on the other hand, does not confirm these convergence tendencies. On the contrary, opinion diversity often remains stable and can actually increase over time.

In this book, we have sought to overcome this conflict. In particular, our goal was to develop theories of social-influence dynamics which are capable explain two specific phenomena. First, following Abelson's lead, we looked for an explanation for *bimodal* opinion distributions. In other words, we sought to explain the parallel persistence of several (at least two) distinct but internally homogeneous subpopulations of actors. Explaining opinion clustering is important because there is empirical research showing that opinions tend to cluster for example within geographical regions (Glaeser and Ward 2006), sociodemographic groups (Mark 2003) and online communities (Lazer et al. 2009). We therefore reformulated Abelson's clustering puzzle as the first research question of this book:

Research Question 1:

How can the persistence of clustering despite social influence be explained?

Existing social influence models fail to generate another phenomenon. As Figure I.2 shows, bounded-confidence models may be able to generate clusters, but the distance between the clusters remains constant over time. There is, however, empirical evidence showing that opinions may polarize in the sense that subgroups with increasingly distinct opinions form. Polarization tendencies have been found in studies on the American public opinion on issues such as abortion, sexual morality and the war in Iraq (DiMaggio, Evans and Bryson 1996; Evans 2003; Fiorina and Abrams 2008; Fischer and Mattson 2009; Levendusky 2009). Furthermore, there is research on conflicts in work teams showing that task and emotional conflicts can intensify over time (Early and Mosakowski 2000). The fact that standard social influence models fail to explain polarization leads to the second research question:

Research Question 2:

How can the polarization of opinions despite social influence be explained?

I.3. Our answers to the research questions

Compared to the three examples that we introduced at the beginning, the existing social influence models appear to be very abstract. For instance, the models ignore possible demographic differences between individuals (see Example 1). What is also neglected is that social-influence may work differently for high-school students (Example 3) than it does for adults (Example 2). High abstraction is a desirable feature of a theory for two reasons. First, abstract theories have more analytical power in the sense that they can be applied to a wider range of empirical realms. For instance, a theory which is perfectly tailored to the opinion dynamics of survivors of a plane crash but can not be applied to other scenarios is not very informative. Furthermore, the more specific a theory is the fewer real-life situations it can be applied to. This makes it difficult to empirically test theories that are tailored to very specific phenomena (Popper 1972; Popper 1959). The second advantage of abstraction is that abstract theories make fewer assumptions. This makes theories simple and relatively easy to understand.

The existing social influence models are analytically powerful and they are simple, but they are also empirically inaccurate. We have shown, for instance, that these models fail to predict opinion polarization and opinion clustering. To overcome these shortcomings, we will need to forgo some of the analytical power and simplicity of the existing social-influence models and add new ingredients to the models. In doing so, we will follow the principles of Lindenberg's method *of decreasing abstraction* (1992). According to this method, one gradually decreases abstraction by adding new assumptions in a step-by-step process. Thus, instead of simultaneously adding several assumptions which potentially help to overcome empirical inaccuracy, the researcher adds one assumption after the other. After having included a new ingredient, one first tests whether the model changes do indeed add empirical accuracy. In other words, one tests whether including the new assumption helps in overcoming the shortcomings of the original model. This is central because adding assumptions which fail to increase the quality of the theory's predictions is a waste of analytical power and theoretical simplicity.

Furthermore, it is sensible to include new ingredients in the original theories independently one by one. This is because two ingredients may affect the predictions of a

theory in a similar way. In this book, for instance, we discuss two different mechanisms, both of which can generate opinion polarization. However, we do not include both of these mechanisms in one model. To keep the number of assumptions in our theories low, we developed two competing models and studied under what conditions the two models predict polarization. We then identified conditions where the two models imply different predictions. These different predictions can then be tested empirically so as to test under what conditions each model implies more accurate predictions.

Finally, we want to point out that we mainly included *individual-level mechanisms* in the existing models. Individual-level mechanisms make assumptions about the behavior of the individuals who constitute the studied population. The social-influence assumption is an example of such a mechanism. In contrast, *structural factors* point to characteristics of the population. For example, it has been argued that opinion dynamics are affected by the existence of extreme opinion leaders (Hegselmann and Krause 2002), competing political parties (Fiorina and Abrams 2008) or media (Watts and Dodds 2007).

We do not question the fact that structural factors can have a significant impact on opinion dynamics. However, we do argue that individual-level mechanisms should be included in a theoretical model before structural factors are considered. This allows for testing whether a given research question can be answered even without including structural factors. For instance, can one explain polarization even in the absence of opinion leaders and political parties, or are certain structural factors necessary conditions for opinion polarization?

A second reason why we focused on individual-level mechanisms is that the effects of structural factors can crucially depend on which individual-level mechanisms are included in a model. In this book, for example, we study the effect of demographic diversity (a structural factor) on opinion polarization with two different models of individual behavior. In Chapters II and IV, we show that the two models have different implications concerning the conditions for this effect. To be able to understand why the two models imply different predictions, it is important to understand the dynamics that each individual-level mechanism generates. This is easier when structural factors are excluded.

In this book, we discuss four ingredients which we have added to the existing social-influence framework. We included three different individual-level mechanisms and one structural factor. In each chapter, we focused on one or two of these ingredients and

studied whether they help answer the research questions. In the following, we will sketch how each ingredient might help in answering the two research questions. However, each of the ingredients has shortcomings, too. We will address these shortcomings in Section I.6.

I.3.1. Negative influence

Classical social-influence models are based on two central assumptions. First, it is assumed that individuals tend to interact with others who hold similar opinions (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002). Second, during interaction individuals influence each others' opinions in such a way as to become more similar (Abelson 1964; French 1956). In their review of the social-influence literature, Mason, Conrey and Smith (2007) recently recommended adding the negative counterparts of these two social mechanisms. In short, this suggests adding two further assumptions. First, individuals with very distinct opinions may not only refuse to interact but may actually dislike each other (Byrne, Clore and Smeaton 1986; Chen and Kenrick 2002; Pilkington and Lydon 1997; Rosenbaum 1986a; Rosenbaum 1986b; Smeaton, Byrne and Murnen 1989). Second, individuals may tend to increase the opinion distance with the disliked others. This notion has been called *negative influence* and has recently been included in social-influence models (Baldassarri and Bearman 2007; Jager and Amblard 2005; Macy et al. 2003; Mark 2003; Salzarulo 2006).

Negative influence could lead to polarization and might thus offer an answer to our second research question. How does it do so? Consider a population where the opinions are normally distributed at the outset. This population is comprised mainly of moderate actors but also a few extremists. The extremists from the opposing poles are highly dissimilar and will influence each other negatively. If this negative influence is strong enough, the extremists may thus become even more extreme. This can have consequences for the moderate actors because they are influenced positively by those extremists who are closer to them. When some moderates become more extreme, this can start a bandwagon effect in which the number of moderates decreases gradually until the population is split up into two maximally dissimilar groups.

I.3.2. Persuasive Arguments

The classical social-influence models assume that actors adopt opinions similar to the *average* opinions of their interaction partners. However, research on the outcomes of group discussions suggests that this assumption may be too simple (Myers 1982; Vinokur and Burnstein 1978). During interaction, individuals not only inform each other about their

opinion, but they also talk about the arguments which they base their opinions on. Furthermore, it is possible that two actors hold identical opinions but base them on different arguments. When they interact and exchange arguments, they will provide each other with new reasons for their opinions. This could intensify their opinions in the sense that they develop more extreme views.

In combination with the selection of similar interaction partners, this elaboration of the influence process could explain opinion polarization. When individuals mainly interact with others who hold similar opinions they mostly exchange arguments which will intensify each others' opinions. As a consequence, also actors with rather moderate opinions would tend to become more extreme over time. This could aggregate to polarization.

For example, such a process might lead to opinion polarization during the Oscar parties (see Example 2). After the award ceremony ends, people will decide to which party they will go to. This decision will likely correlate to their opinion about the Academy decision. The friends of the winner and his colleagues will be very happy with the decision and will want to celebrate with the winner. Those, however, who do not have a close relationship with the winner but have a close relationship with one of the other nominees will likely hold more negative opinions. These people may not want to attend the party of the winner but may prefer to meet the losers to cheer them up. The consequence of these decisions is that the participants in a specific party will likely hold similar opinions. When people then talk about the decision of the Academy, they will likely provide each another with new arguments which support their initial opinions. As a consequence, initial opinion differences might intensify during the Oscar night.

I.3.3. Striving for uniqueness

Example 3 is concerned with the opinions towards rap music in a class of high-school students. These students are adolescents and, thus, are in a phase where they are very open to influence from peers. Hence, one would expect that after a few weeks all classmates will have adopted the opinion of the rap fan. On the other hand, adolescents also seek to define a *unique* identity for themselves. As a consequence, they might want to disagree with the majority opinion in their class. This suggests that students might want to deviate from an emergent opinion consensus and that consensus will be very unstable. We argue that the interplay of peer influence and striving for uniqueness can lead to development of several homogeneous subgroups with different opinions. Such constellations will be relatively stable because students will feel unique as they perceive

opinion differences with many classmates. At the same time, there are also classmates with similar opinions. This will satisfy the students' need to be similar to peers.

Prominent sociological theories of social differentiation (Durkheim 1973; Durkheim 1982 [1895]; Simmel 2004 (1858); Turner 1995) suggest that striving for uniqueness is a strong force in modern societies. These theories argue that population growth and the development of new technologies initiate competition for resources. This, in turn, forces individuals to distinguish themselves from the mass. As an example, individuals specialize in their occupations in order to differentiate themselves from competitors. But in addition to this functional differentiation, individuals also want to differentiate themselves in terms of cultural dimensions and develop individual opinions and values. This idea is supported by recent research on the need for uniqueness (Imhoff and Erb 2009; Maslach, Stapp and Santee 1985; Snyder and Fromkin 1980) which found that individuals tend to deviate from behavioral regularities.

Moreover, Brewer's optimal distinctiveness theory (Brewer 1991) and Social Identity Theory (Tajfel and Turner 1986) hold that there is a "fundamental tension between human needs for validation and similarity (on the one hand) and a countervailing need for uniqueness and individuation (on the other)"(Brewer 1991: 477). However, even though these theories capture opinion dynamics, research along these approaches has mainly focused on how humans self-categorize. In other words, attributes such as opinions are often assumed to remain unchanged when individuals feel too similar or too dissimilar to others. Instead, the research focuses on which set of attributes humans consider important when they define their identity. We went beyond these approaches and explicitly included the possibility that individuals may adjust their opinions in order to decrease tension between the needs for similarity and uniqueness.

We want to point out that the striving for uniqueness differs crucially from negative influence. Negative influence presumes that individuals want to increase existing differences from *dissimilar* others. This motivation is stronger the more *dis*similar the respective other person is. Contrary to this, the striving for uniqueness implies that individuals are trying to create differences from those they consider to be too *similar*. As soon as this similarity has decreased, the motivation to individualize will also decrease.

I.3.4. A structural factor: demographic faultlines

In Example 1, four passengers survived a plane crash and then try to escape from a jungle. There are two males and two females. Two are black and two are white. Classical

sociological theorizing on social differentiation in modern societies would suggest that demographic differences such as these can influence the outcomes of group processes (Bourdieu 1984[1979]; Elias 1969[1939]; Simmel 1957). These theories hold that individuals want to set themselves apart from distinct social categories. To signal that they are different from distinct groups, individuals strategically reject cultural tastes, attitudes and behaviors which they consider typical for these groups or which have been adopted by outgroup members (Berger and Heath 2008; Bryson 1996).

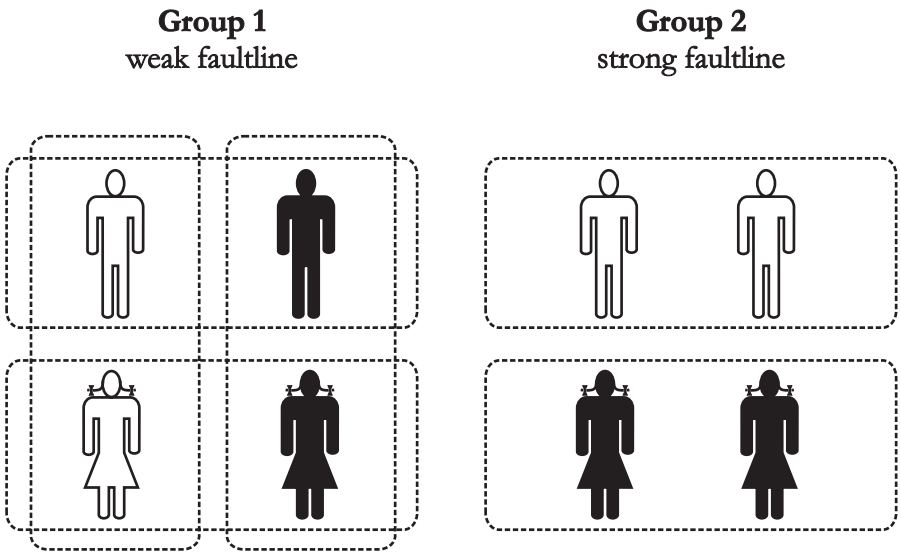
Along the same line of reasoning, socio-psychological theories of intergroup relations propose that humans want to accentuate differences between distinct social categories (Brewer 1991; Tajfel and Turner 1986; Turner 1987). These theories hold that as an ingroup becomes psychologically salient, individuals will adjust their opinions in such a way as to conform to the ingroup's stereotypical position (Turner 1987). This position will be similar to the average opinion of the ingroup members. In contrast, when an outgroup is salient too, the stereotypical position will shift in a way so as to "distinguish the ingroup clearly from relevant other groups and thus *maximize* differences between ingroups and outgroups" (Hogg, Turner and Davidson 1990 p. 80). To summarize, theories of social differentiation and social-categorization theories hold that demographic differences can trigger negative influences. This might lead to polarized opinions in the sense that, for instance, the two black survivors will want to go in the opposite direction from the two whites.

In the group of survivors, demographic diversity is very high in the sense that the group can be separated along two dimensions, gender and skin color. Intuitively, one would expect opinion polarization to be very likely in this group. However, Lau and Murnighan (1998; 2005) proposed that the effect of demographic diversity may decisively depend on the way demographic attributes are distributed across the two dimensions (see also Colson 1954; Evans-Pritchard 1939; Flap 1988; Galtung 1966; Lijphart 1977; Ross 1920; Simmel 1922 (1908)). They claim that demographic diversity causes opinion polarization only when the distribution of demographic attributes generates a *strong faultline*. "Group faultlines increase in strength as more attributes are highly correlated, reducing the number and increasing the homogeneity of resulting subgroups. In contrast, faultlines are weakest when attributes are not aligned and multiple subgroups can form" (Lau and Murnighan 1998: 328).

Figure I.3 illustrates Lau and Murnighan's faultline concept. The figure shows two groups of four actors each. Similar to the plane-crash survivors, both groups consist of two

males and two females, on the one hand, and of two blacks and two whites, on the other. Thus, both groups are characterized by a high demographic diversity. However, the two groups differ in faultline strength. In Group 1, the two demographic dimensions are not aligned. The boxes indicate that each group member shares one demographic attribute with two other members of the group. This similarity might prevent negative-influence tendencies and therefore hamper opinion polarization. In Group 2, however, there are two white males and two black females. Thus, pairs of group members are either perfectly similar or perfectly dissimilar. The clear-cut divide between the two subgroups might cause negative influence tendencies and thus lead to opinion polarization.

Figure I.3: Two groups with identical diversity but different faultline strength



I.4. Our methodological approach: ABC modeling

In each chapter of this book, we added a new ingredient to the existing social-influence models and demonstrated how each of them helps to answers our research questions. Furthermore, in each chapter we pointed out *counter intuitive* and unexpected implications of our new models. For example, we claimed in chapter III that opinion polarization can emerge even though individuals do not seek to increase opinion differences with others. More precisely, we proposed that perfectly homogeneous populations can fall apart into subgroups which develop increasingly distant opinions even though individuals do not seek to increase opinion differences. This proposition contradicts intuition as well as standard theories of intergroup processes (Brewer 1991; Tajfel and Turner 1986; Turner 1987) and classical theories of social differentiation

(Bourdieu 1984[1979]; Elias 1969[1939]; Simmel 1957). Contrary to our approach, these theories predict that opinion differences will increase only in settings where there are already initially salient differences between subgroups of individuals. Furthermore, these theories hinge on the assumption that individuals seek to increase opinion differences with dissimilar others (negative influence). To corroborate our provocative propositions, we developed formal models of our social-influence theories. With these formal models we demonstrated that our propositions follow consistently from the assumptions we made.

In this book we developed so-called *agent-based computational models* (ABC models). In line with social-influence theories, ABC models (Bonabeau 2002; Macy and Willer 2002; Macy and Flache 2009; Smith and Conrey 2007) study interactions between multiple actors, called agents, who react to influences they receive from one another.

To develop an ABC model of a given theory, the theory's assumptions about the individuals' behavior are translated into a formal language and implemented in a computer program. Starting form given initial conditions (e.g. a certain initial opinion distribution, etc.), the behavior of each agent (which follows from the theoretical assumptions) is then calculated by the computer program. Dynamics are broken down into multiple consecutive events. For example, in ABC models of social-influence theories, the computer program typically picks an agent and updates this agent's opinion. In the subsequent event, another agent's opinion is updated. In this way, opinion changes at the previous event(s) are taken into account. In most cases, events are iterated until dynamics reach equilibrium.

The advantage of ABC models is that this method allows studying dynamics in big populations of interdependent individuals and including multiple nonlinear and probabilistic assumptions. For the classical social-influence models, central propositions have been proved using analytical methods (Abelson 1964; Berger 1981; DeGroot 1974). For instance, it has been demonstrated that opinions always converge to perfect uniformity unless no subset of actors is perfectly separated from the population. The Bounded Confidence model (Deffuant, Huet and Amblard 2005; Hegselmann and Krause 2002), however, is already so complex that analytical results are available only for a limited part of the parameter space (e.g. for small groups and large and homogeneous confidence intervals) (for an overview of available analytical results see Castellano, Fortunato and Loreto 2009; Lorenz 2005; Lorenz 2007). In this book, we added further ingredients to the existing models and thereby increased their complexity even more. This suggests that analytical results are hard to obtain for the propositions presented in this book. Accordingly, we followed recent work on extensions of conventional social-influence

models (see e.g. Baldassarri and Bearman 2007; Mark 2003) and employed computational agent-based modeling.

I.5. Overview of the chapters

In each chapter of this book, we discussed a potential answer to our research questions. In doing so, we always took the existing social-influence models as a baseline and added new assumptions. Thus, all models started from the assumption that individuals tend to interact with those members of their group who hold similar opinions. Furthermore, during interaction, individuals exert influence on each others' opinions.

In **Chapter II**, we added the negative-influence assumption. Recently, several publications have taken this step and have demonstrated that social-influence models can generate opinion polarization when the negative-influence assumption is included (Baldassarri and Bearman 2007; Jager and Amblard 2005; Macy et al. 2003; Mark 2003; Salzarulo 2006). We have contributed to this finding by including a second ingredient, demographic attributes. To be precise, we studied the effects of demographic faultline strength on the opinion dynamics in groups.

In Chapter II, we applied our social-influence model to work teams. Why? Because empirical research suggests that opinion dynamics in a team can affect the team's work performance. In particular, research has confirmed that consensus on opinions decreases emotional and work-related conflicts in the team (Harrison et al. 2002; Jehn 1994; Jehn, Northcraft and Neale 1999). Fewer conflicts, in turn, increase team performance and individual work satisfaction (Jehn 1994; Jehn 1995). However, work teams that perform non-routine tasks might also profit from some disagreement because this can promote inspiring discussions and critical evaluation of problems and decision options (Jehn 1995; Jehn 1997; van Knippenberg, De Dreu and Homan 2004). It has been found that there might be an "optimal level of task conflict in nonroutine-task groups" (Jehn 1995: 275). This optimum is reached when there is enough disagreement to trigger an inspiring discussion, while, at the same time, there is also enough agreement to prevent conflicts.

Furthermore, in an increasingly globalized economy, the *demographic* composition of work groups has become a central issue for organizations (Bowers, Pharmer and Salas 2000; Milliken and Martins 1996; Pelled 1996; Stewart 2006; van Knippenberg and Schippers 2007; Webber and Donahue 2001; Williams and O'Reilly 1998). Diversity in dimensions such as ethnic background, religion and gender can be beneficial for

organizations, because it broadens the social and human capital that teams can use to fulfill their tasks. At the same time, demographic differences can cause disagreement and tensions which threaten performance. Out of these considerations, the very practical question emerges: Under what conditions will demographically-diverse work teams manage to overcome disagreement and be able to make use of their advantages?

Our analyses in Chapter II suggested that work teams with high demographic diversity but a weak faultline will not suffer from polarization. However, the stronger the faultline the more likely opinions will polarize. Furthermore, we demonstrated that the model highlights a new structural condition that may give managers of work teams a handle for tempering faultline effects. We argued that managers might do well to manipulate the *timing of contacts* in work teams. It follows from our model that newly formed teams with strong demographic faultlines very likely polarize when group members freely interact with each other. Because of the high demographic dissimilarity, team members reject the opinions of dissimilar colleagues and develop extreme opinions. However, we showed that opinion polarization is less likely when teams are first separated into demographically homogeneous subgroups and are merged only later in the opinion-formation process. Because of the demographic similarity between members of subgroups, it is very unlikely that they will influence each other negatively. Hence, subgroups will likely develop local consensus on moderate opinions. When subgroups are merged after they have found local consensus, opinion similarity will act as a bridge over the demographic faultline and lead to further convergence.

Besides the obvious practical implications of this intervention, the timing of contacts hypothesis also challenges a prominent theory of intergroup relations; contact theory (Allport 1954). Contact theory holds that contact between members of different social categories will improve interpersonal relations. In contrast, our model implies that contact between dissimilar group members in the early phase of the opinion formation process might intensify initial opinion differences and impair interpersonal relationships.

In **Chapter III**, we developed an alternative explanation for opinion polarization. We started by pointing out that prominent social-psychological theories of group processes (Brewer 1991; Turner 1987) hinge on the assumption that individuals seek to increase opinion differences with dissimilar others. However, empirical research on negative influence provides very little evidence for this mechanism (e.g. Krizan and Baron 2007), suggesting that negative influence shapes opinion dynamics only in a restricted number of settings. In Chapter III, we proposed an explanation for opinion polarization which does

not hinge on the negative-influence mechanism. Our counter-intuitive proposition is that perfectly homogeneous populations can also split up into subgroups with maximally distinct opinions, even when individuals do not strive for any opinion differences. We developed a new social-influence model that includes the persuasive-argument mechanism and demonstrated that this model can generate opinion polarization. In addition, our analyses revealed that a *strong* individual tendency to interact with those group members who hold similar opinions is a central precondition for polarization.

Chapter III also reported the results of a laboratory experiment (N=96) which we designed to test the new theory. In this experiment, participants discussed an opinion in a computer network. We created a setting which rules out negative-influence tendencies and tested under what conditions there was opinion polarization. It turned out that there was no polarization when participants only informed each other about their opinions. However, in those experimental conditions where participants exchanged arguments, there was significant polarization. Furthermore, we included a condition where participants exchanged both arguments and opinions. In this condition, we also found significant opinion polarization. In the end, the experiment confirmed the hypothesis that the selection of similar interaction partners is an important precondition of polarization.

In **Chapter IV** we applied the model which we have developed in Chapter III to work teams. As in Chapter II, but now with the new model from Chapter III, we studied the effects of faultline strength on opinion polarization. We showed that also the argument-exchange model supports the faultline hypothesis. However, the new model points to several conditions for this effect which previous contributions have overlooked. First, even with a very strong faultline, opinions will only polarize in those groups where individuals tend to select similar interaction partners. Second, polarization is more likely, the stronger the opinions and demographic attributes in a team are correlated initially, that is, prior to interaction between the group members. Furthermore, the new model implies that the short-term effects of demographic faultlines differ crucially from their long-term effects. Groups where demographic attributes are not perfectly correlated will eventually arrive at consensus, even though they might suffer from polarization in the short run. Counter-intuitively, the model also implies that the convergence process is faster, the *stronger* the demographic faultline is. In other words, even though teams with strong faultlines might suffer from opinion polarization in the short run, they might arrive at consensus faster than teams with a weak faultline.

In **Chapter V**, we focused on the uniqueness mechanism and tested whether it helps explain opinion clustering. We took the existing social-influence models as the baseline and included the assumption that individual opinions are also shaped by the need to hold a unique opinion. In an initial step, we included striving for uniqueness as random individual opinion perturbations (white noise). However, we showed that this fails to explain opinion clustering. Weak opinion noise triggers social-influence cascades that lead to consensus. Increasing noise implies rampant individualism rather than clustering.

We presented a new solution to the clustering problem and showed that the new model does generate clustering. The key element of our model is an adaptive kind of noise. We included the factor that the individual striving for uniqueness increases when many members of the population hold similar opinions. However, when actors agree with only a few others, then they are sufficiently unique and do not seek to individualize. In a computational experiment, we identified conditions under which populations develop clusters with diversity between and consensus within clusters. Once they have developed, the clusters are temporarily stable but may later merge. Merging, however, triggers the development of new clusters. In this way, clustering is a stable outcome. Paradoxically, the new model's predictions are not only robust to noise, but noise is the central mechanism that causes cluster formation.

Table I.1 summarizes which ingredient is studied in each of the four chapters.

Table I.1: Ingredients added to the existing social-influence models

	Chapter II	Chapter III	Chapter IV	Chapter V
Individual level mechanism				
Negative influence	×			
Persuasive arguments		×	×	
Striving for uniqueness				×
Social-structural condition				
Demographic faultlines	×		×	

I.6. What have we learned and where do we go from here?

In the following, we will summarize the main strengths and weaknesses of each of our four answers to the research questions. We will also point towards elements for future research.

I.6.1. Negative influence

Recently, several researchers have integrated the negative-influence assumption into the existing social-influence framework (Baldassarri and Bearman 2007; Jager and Amblard 2005; Macy et al. 2003; Salzarulo 2006). In line with their publications, we showed in Chapter II that this approach helps to answer one of our research questions: The new model can explain opinion polarization.

The new model, however, does have two main shortcomings. First, the negative-influence model fails to generate opinion clustering. To be more precise, the model is able to explain the development of distinct clusters. However, these clusters will always hold maximally extreme opinions. This is because members of two emergent clusters will either exert positive influence on each other and find consensus, or the members will influence each other negatively and will become maximally dissimilar. In other words, the model predicts either minimal diversity (consensus) or maximal diversity (polarization). Only in a very few scenarios does the model generate opinion distributions with opinion diversity and, at the same time, some moderate agents. Such opinion distributions consist of maximally extreme clusters and moderate agents. The influences which the extremists exert on the moderate agents balance each other out and the moderate agents do not change their opinions anymore. However, this equilibrium is fragile and can be destroyed by minute opinion changes on the part of a single agent. For instance, if one of the extremists happens to develop a slightly less extreme opinion, then the moderate agents are "pulled" less strongly towards this agent. As a consequence, the influence from the opposite extreme cluster on the moderates will exceed the influence from their opponents. Hence, the moderate agents will adjust their opinions and will become extreme, too.

In sum, the negative influence model generates either minimal opinion diversity (consensus) or maximum opinion diversity (polarization). We consider it a weakness of this model that it fails to generate opinion distributions that fall between these two extremes.

The second and even more problematic shortcoming of the negative-influence model is discussed in Chapter III. Empirical research on negative influence has led to very mixed results. There are publications (Berscheid 1966; Mazen and Leventhal 1972; Sampson and Insko 1964; Schwartz and Ames 1977; van Knippenberg and Wilke 1988) which claim that they found support for negative influence. However, in Chapter III we pointed out methodological shortcomings to these studies. In addition, there are also studies which did

not confirm the negative-influence assumption at all (Hogg, Turner and Davidson 1990; Krizan and Baron 2007; Lemaine 1975).

We do not claim that the negative-influence assumption should be considered falsified and excluded from social-influence models. However, we do need to be careful when we apply the negative-influence model to specific real-life settings. We cannot be confident about the model's predictions when we are not sure that negative influence might play a role in a specific setting. We therefore feel that future empirical research is needed to identify the conditions for negative-influence tendencies. For instance, individuals may be influenced negatively by dissimilar others only when they perceive significant opinion differences in the overall population. In other words, negative influence may be generated only once other factors have set opinion polarization in motion. Considering the methodological difficulties of disentangling negative from positive influence (see Chapter III), we suggest developing laboratory experiments to test hypotheses about the conditions of negative-influence.

In addition, the timing-of-contacts proposition points to an indirect test of the negative-influence assumption. In Chapter III, we demonstrated that opinion polarization is less likely when dissimilar agents are first separated and brought in contact only later in the influence process. This proposition critically hinges on the negative-influence assumption. This, in turn, implies that empirical support for the timing-of-contacts proposition indirectly confirms the hypothesis that negative influence shapes opinion dynamics. Hence, timing-of-contacts would appear to be an interesting independent variable for future experimental research.

I.6.2. Persuasive arguments

In Chapters III and IV, we showed that including the persuasive-argument principle in the social-influence framework helps to explain opinion polarization. What is more, our new model is able to generate opinion polarization even when there are no opinion differences between the agents at the outset. A central strength of the persuasive-argument mechanisms is that empirical research has consistently confirmed that argument exchange shapes opinion influence. More to the point, the exchange of arguments can intensify opinions in the sense that individuals develop more extreme views when they learn new arguments which support their opinions (Myers 1982).

However, this model also has weaknesses. First, also the persuasive-argument model is unable to explain clustering. Like the negative-influence model, opinion variance is either

minimal or maximal at the end of the influence process. According to the persuasive-argument model, non-extreme clusters will not be stable because their members will interact with members of other clusters. In this way, they will likely be exposed to arguments which will change their opinion. Such arguments will then spread in their cluster and influence the remaining cluster members.

A second weakness of the persuasive-argument model is that opinion polarization is very fragile. In the current version of the model, opinion polarization is stable because members of distinct groups refuse to exchange arguments. If one includes, however, a very small interaction probability for pairs of agents that would refuse to interact otherwise, there will be argument exchange between the subgroups. As a consequence, opinions will converge in the long run. This, however, is not in line with the results from empirical research. Empirical studies on work teams in organizations suggest that, even though team members interact frequently, teams sometimes fail to overcome opinion differences (Early and Mosakowski 2000). The current model fails to explain this finding.

Nevertheless, the persuasive-argument mechanism has found considerable support in empirical research and therefore offers a promising solution to the polarization problem. Most of this empirical research, however, has been performed in the laboratory (Isenberg 1986; Myers 1982). More research outside the laboratory is needed.

Online discussion sites appear to be a promising research field. It has been argued that people tend to visit those discussion sites where they expect to meet people who hold similar opinions (Sunstein 2008). As an example, there are lively discussions about American healthcare reform on the web pages of the Tea Party movement (Tea Party Patriots 2010) and its political counterpart the Coffee Party movement (The Coffee Party USA 2010). If the supporters of a new healthcare system discuss it mainly on the Coffee Party's web pages and if its opponents mainly discuss it on the Tea Party pages, then we might observe opinion polarization during the discussions. One advantage of Internet studies is that they provide the researcher with very detailed information about the opinions of the discussants, their discussion partners and the arguments they use. Furthermore, online discussions can be tracked over very long periods. This allows one to study under what conditions polarization will be stable in the long run. This information might help develop a new persuasive-argument model which can explain stable group splits.

I.6.3. Striving for uniqueness

In Chapter V, we introduced the assumption that individuals strive for uniqueness in the sense that they want to hold opinions that are shared by relatively few others. When too many others hold similar opinions, individuals are motivated to change their opinions (Imhoff and Erb 2009; Maslach, Stapp and Santee 1985; Snyder and Fromkin 1980). We demonstrated that our new model is able explain that opinion clusters emerge and remain temporarily stable. Clusters can merge but the resulting group will not be stable and will split up into clusters again. In this way, clustering is a stable phenomenon.

Our new model, however, offers a somewhat problematic solution to the polarization problem. To be more precise, our model can generate polarization. When clusters have formed, they independently perform a random walk through the opinion space. As a consequence, clusters can happen to develop increasingly distant opinions. This constitutes opinion polarization and the model thus offers a solution to the polarization problem. Still, such scenarios are *random* events. Psychological research on discussion groups, however, suggests that under certain conditions discussion groups *systematically* develop increasingly extreme opinions (Isenberg 1986). What is more, the laboratory experiment which we presented in Chapter III has found statistically significant polarization tendencies. Such systematic polarization tendencies cannot be explained with the new model.

Nevertheless, empirical research supports the assumption that individuals strive for uniqueness (Imhoff and Erb 2009; Maslach, Stapp and Santee 1985; Snyder and Fromkin 1980). Future empirical research is needed to test whether this striving affects opinion dynamics in groups. For this purpose, online discussion sites appear to be a good setting. In contrast to group discussion experiments (Johnson and Johnson 1982), for example, users of online discussion sites form *large* communities. When a consensus emerges, users may thus feel that there are too many others around who hold opinions similar to the ones they do. As a consequence, people might feel less unique and change their opinion. In standard experimental discussion groups, such dynamics may be less likely because of the small size of the discussion groups (usually N=4).

I.6.4. Demographic faultlines

In Chapters II and IV we studied the effects of demographic faultlines on opinion dynamics. We showed that the negative-influence model and our model with persuasive

arguments both predict that opinion polarization is more likely the stronger the demographic faultline is.

Recent empirical tests of the faultline hypothesis have led to inconsistent findings. Some studies found that strong faultline groups tend to suffer from little integration and from emotional conflicts between the team members (Early and Mosakowski 2000; Li and Hambrick 2005; Rico et al. 2007). However other studies found that faultline strength inhibits conflicts and improves psychological safety, job satisfaction and learning behavior in work teams (Gibson and Vermeulen 2003; Hart and Van Vugt 2006; Lau and Murnighan 2005). Thatcher et al. (2003) found curvilinear effects for faultline strength. They report more conflicts and lower performance in teams with weak and very strong faultlines than in teams with moderate faultline strength.

In trying to explain these inconsistent findings, researchers have included moderating variables in their statistical models. It has been shown that faultline effects increase in strength as the distance[2] between faultline subgroups increases (Bezrukova et al. 2009; Molleman 2005). High team identification was shown to prevent negative faultline effects (Bezrukova et al. 2009). Furthermore, team autonomy seems to worsen the effect of faultlines on cohesion and integration (Molleman and Slomp 2006; Rico et al. 2007).

In Chapter III, we pointed out three additional variables that might moderate the effects of faultline strength on opinion polarization and conflicts in work teams. First, the persuasive-argument model implies that strong demographic faultlines breed opinion polarization only when team members tend to interact with similar others. Second, opinions and demographic attributes already need to be correlated at the beginning of the social-influence process. Third, time may play a crucial role. In particular, the model implies that teams with a strong (but not maximally strong) faultline will overcome polarization in the long run. What is more, even though strong faultline teams might suffer from polarization in the short term, these teams might end up reaching consensus faster than weak faultline teams. This would suggest that results of empirical studies may critically depend on when opinion differences and conflicts are measured. Future empirical work should therefore focus on longitudinal data.

[2] For example, two 30-year-old team members differ more from two team members aged 50 than from two team members aged 40.

I.6.5. Achievements and prospects in a nutshell

In this book, we sought to explain two phenomena that existing social-influence models fail to explain; opinion clustering and opinion polarization. Our strategy was to start from the existing social-influence framework and to include new ingredients. We developed several formal models and studied whether and under which conditions each of them are able to generate clustering and polarization.

We demonstrated that each of the models provides an answer to one of our research questions. The models with negative influence and persuasive arguments are able to explain opinion polarization. The uniqueness model is able to explain opinion clustering. However, none of the three models offers a satisfactory explanation for both clustering and polarization.

Future work is needed to test the new models empirically. First, empirical research should test the assumptions we have added to the existing models. This is of central importance for the negative-influence assumption. Second, empirical research should test the new predictions of our models. Based on the results of these empirical tests, one will have to decide which of the models is most promising and is capable of forming a reliable basis for future modeling work. Alternatively, two of our ingredients could be integrated into one model.

In this book, we offered new solutions for the hitherto unresolved puzzle of opinion diversity despite social influence. Much work remains to be done, but we believe that the theories put forward in this book open up directions for fruitful future research. Moreover, they suggest potential new approaches for the management of opinion dynamics in groups that may help practitioners to avoid polarization or, if necessary, sustain a healthy diversity of opinions.