Everyday Diplomacy

Roos, Carla

DOI:
10.33612/diss.230455324

Chapter 2

The social dynamics approach to mediated communication

Roos, C.A., Koudenburg, N., & Postmes, T. (2022). *The social dynamics approach to mediated communication* [Manuscript in preparation]. Department of Social Psychology, University of Groningen.

We thank professor Emiel Krahmer and professor Marjolijn Antheunis for their valuable input. We further thank Jannis Kreienkamp for designing Figure 1.

Abstract

Since the early 1980's, the use of mediating devices for communication has witnessed an extraordinary growth. Not only quantitatively, but also qualitatively the mediated communication technologies underwent enormous developments (e.g., from text-only to video and voice messages). Not surprisingly, empirical research in this field is also rapidly evolving. Theoretical development is lagging behind, however, with all major theories dating back to around the turn of the century. As a consequence, the current mediated communication literature is very much focused on how the mediating technology, or its affordances, changes individual cognitions and behaviors – a mechanical approach. But these mechanical views have difficulty explaining contradictory findings, for instance in the domain of online polarization. In this paper we integrate the mechanical with a social constructivist approach by taking into account the collective and collaborative nature of interaction while at the same time acknowledging a technological reality. This results in a new theoretical and methodological framework for studying mediated communication: The Social Dynamics Approach to Mediated Communication (SoDA). We believe this approach provides a better understanding of social dynamics within interactions *and* through that also gives insight into the differences we observe between communication media. We illustrate how this new framework can be used in mediated communication research, what new insights it might bring, and how it might help towards solving contradictions in the literature. We close the paper with implications and directions for future research.

The social dynamics approach to mediated communication

Abundant research and thinking in the domain of computer-mediated communication (CMC) contains assumptions about how properties of technology determine usages. For example, the literature suggests that conversing online can lead to misunderstandings and can easily polarize, because receivers tend to interpret online messages more negatively than the sender intended (the "negativity effect", Byron, 2008; Sillars & Zorn, 2021), or because online users would feel that they can be aggressive and hostile without consequences ("toxic disinhibition", Suler, 2004; Stuart & Scott, 2021), or because online people can readily find so much information that they self-select sources and contacts that confirm their worldviews and strengthen prior convictions ("echo chambers", e.g., Cinelli et al., 2021).

But the literature also shows many exceptions to such straightforward accounts of online polarization: online interactions do not *inevitably* produce polarization. For example, interaction partners can be very sensitive to social norms online (Huang & Li, 2016; Postmes et al., 1998) and they need not come into conflict either; online technologies are in fact very suited to foster intimacy (Walther, 1998; Walther & Whitty, 2021). It is common to explain such apparent contradictions by assuming that the possibilities of CMC offer users the ability to display "bad" or "good" behavior depending on their social motives. That is, the possibilities of technology would interact with what users want or need.

Irrespective of whether theories predict the usage of CMC to have uniformly negative or positive effects for human behavior, or they predict that motives interact with the possibilities offered by technology, these theories all contain mechanical cause-effect explanations of how human behavior is changed online. These mechanical approaches can be contrasted to social constructivist or structuration approaches which tend to point to the social dynamics that give rise to the emergence of, among others, disagreements or agreements (DeSanctis & Poole, 1994; Giddens, 1984; Searle, 1990, 1995). Here, it is posited that social norms, together with technological factors, shape interaction behaviors, and vice versa: interaction behaviors shape social norms. These approaches are much better equipped to explain the wide variety of sometimes unexpected social practices that can emerge online. But their downside is the mirror image of technological determinism: they are essentially indeterminate, and can therefore leave us with the impression that "anything goes" and "it all depends".

The present paper aims to integrate these two very different approaches. This has benefits both for theorizing and research of media effects. Our attempt to bridge

and combine the mechanical and constructivist approaches resulted in a new theoretical and methodological framework for studying mediated communication: The Social Dynamics Approach to Mediated Communication (SoDA). In short, we propose that the technical possibilities, personal goals, and social norms within an interaction context are jointly constructed by interaction partners through a dynamical process in which they make sense of their relationship and their collective goals within the interaction. We refer to this interaction partner's idea of who they are together and what they are doing collectively as their *perceived social structure*.

The main reason for developing this new theoretical and methodological framework is that we believe technology is not "neutral" and profoundly changes human behavior. But we can only understand the impact of media on human behavior (i.e., a between-media comparison) by closely studying the processes within media that can explain these between-media differences (i.e., a within-media analysis). By acknowledging this, the SoDA enables us to better understand the processes and outcomes of mediated communication, and helps us explain some apparently contradictory findings across literatures.

The paper will start with a short outline of the history of the CMC literature and a reflection on the current state of the field. Then, we will present our theoretical framework and subsequently show how to translate this to the empirical reality in the form of a methodological framework. After that, we will describe our recent line of research as a concrete example of how the SoDA can be put into practice. We then reflect on the new insights that this line of research gave into three research areas of central interest to the mediated communication literature: media richness, online disinhibition, and online polarization. We end this paper with implications and directions for future research.

The Effect of the Medium

Ever since people started to use computers for communication in the early 1980's, the computer-mediated communication (CMC) field has witnessed rapid and radical developments. In the nineties, all CMC was exclusively text-based, only supported by desktop computers, and mostly used in a business context by only a handful of people. In current day society, that is the 2020's, things look rather different. First, virtually all individuals in developed countries have internet access through one or multiple wearable devices (e.g., smartphones). Almost everyone can be connected always and anywhere. Second, besides the immensely increased quantity of CMC, also its quality or form has thoroughly changed. The mediating devises are "richer" in the sense of allowing the transfer of cues beyond short texts, such as pictures and voice messages but also video calls, and online (social media) platforms enable users to communicate to a public of hundreds or thousands

anonymous others. Following these technological developments, the empirical research in the CMC field has experienced an equally explosive growth. But paradoxically, there has been hardly any theoretical innovation and all major theories date back to before the turn of the century (Mason & Carr, 2021; Yao & Ling, 2020), for example, social information processing theory (Walther, 1992), Media Richness Theory (Daft & Lengel, 1986), Social Presence Theory (Short, Williams, & Christie, 1976), Adaptive Structuration Theory (DeSanctis & Poole, 1994), the Social Identity model of Deindividuation Effects (Postmes et al., 1998).

The CMC literature has been dedicated to mapping and explaining how communication technologies shape users' cognitions and behaviors. In CMC research, the communication medium, or its properties (e.g., anonymity), is the independent variable: a stimulus that has a response. This is then compared to other media or properties, and in this comparison face-to-face (FtF) interactions and their characteristics are seen as a *baseline*. In this way, CMC research emerged from an often implicit set of standards and expectations about what is "normal" in FtF interactions. For example, based on the observation that people tend to use more rude and threatening language (response) when they communicate via online media (stimulus) than when they communicate FtF (Lapidot-Lefler & Barak, 2012; Stuart & Scott, 2021), researchers concluded that online media use, by increasing anonymity, reduces people's social concerns and disinhibits their behavior (Suler, 2004).

Since technologies for mediated communication keep evolving and new ones keep emerging, the once-ubiquitous CMC-FtF comparison has become less widespread. It became more relevant to know about differences between different CMC technologies, such as comparing the degree of romantic attraction after text-based CMC versus videoconferencing encounters (Antheunis et al., 2020), or to learn about the effects of changing the properties of existing CMC technologies, for example, studying the effect of doubling the permissible length of a tweet on uncivility (Jaidka et al., 2019). Although these studies still use the same stimulus-response paradigm in studying technology effects, they offer us ever-more refined insights in properties of technology and their potential effects. We should also note that in many other studies the baseline is implicit. Consider studies measuring the association between internet usage and incidence of depression (Nakagomi et al., 2022), or assessing whether social media consumption is related to online political participation (Yamamoto et al., 2020).

It is clear that the field has moved on immensely, but implicitly or explicitly recent studies still rely on assumptions about richness, presence and so forth which derive from a CMC-FtF comparison: the stimulus-response tradition. This mechanical approach to studying CMC has clearly been very successful. However, as

we shall elaborate further, the stimulus-response paradigm ignores certain key *processes* that also explain the behavioral and cognitive phenomena that this research has investigated.

Changing Field: Processes and Affordances

Recently, in a special issue in one of the defining journals of this field, Journal of Computer-Mediated Communication, two prominent CMC researchers called to dig deeper by focusing attention on the *process* of mediation rather than on the mediating device (the computer and its software) in order to provide input for more enduring theory (Carr, 2020; Flanagin, 2020). In line with this, Carr proposed to rename the field to mediated communication (MC), thus removing the term 'computer'. In the slipstream of quickly changing technologies, and based on earlier approaches (variable-centered approach, Nass & Mason, 1990; mix of attributes perspective, Eveland, 2003; affordances approaches, e.g., Fox & McEwan, 2017), both Carr and Flanagin call for moving focus from the novel technologies themselves, the "object-centered approach" (Flanagin, 2020, p.24), to the (different) processes of communication enabled by these technologies. For this, they stress the central importance of affordances. Affordances are the possibilities that a particular environment offers for behavior (Gibson, 1979), and can in this case be seen as the link between the mediating technology and the outcome. As an example, Flanagin mentions how technological features such as "likes" (Facebook, Instagram), "retweets" (Twitter), and "upvotes" (Reddit) provide the affordance of opinion visibility, which could be used to explain the outcome of political opinion conformity. Carr and Flanagin therefore propose to focus on the realm of possible behaviors within communication environments (affordances; how can people act?), rather than focusing on the effects of the medium.

The affordances literature is broad and diverse, but many researchers agree that whether and how affordances actually affect behavior depends on the user's subjective interpretation of these affordances and the user's goals within a certain interaction context (e.g., Faraj & Azad, 2012; Norman, 1988; Rice et al., 2017). First, the user needs to perceive the affordances offered by the medium in order to be able to use them. For example, if the user does not know that a certain communication medium affords them anonymity, the user will not be able to "use" this affordance and it will not affect their behavior. Second, the way the perceived affordances affect behavior depends on the user's (interpersonal) goals and their assessments of what behavior is needed to accomplish these goals. For example, the anonymity afforded by an online chat environment might be "used" to flame or troll by a user that wants to insult someone without consequences, *or* to present oneself in a very socially desirable way by a user that aims to woo someone. Affordances, defined as

possibilities for behavior, would thus only materialize in the interaction between the user's goals and the medium: affordances are used strategically.

Carr and Flanagin's affordances-based approach suggests a different way of looking at mediated communication: rather than examining the impact of a stimulus environment on behavior and cognition, we need to examine the process by which behavior unfolds in the interaction between user and environment. It thus changes the focus of theorizing from "people reacting" to the context of the medium to "people acting" inside an environment equipped with affordances (both real and perceived) and user's personal and interpersonal desires and goals. So, besides technical constraints, there are *personal* constraints to behavior.

The term affordance stems from an ecological, dynamical perspective on communication (Gibson, 1979). One could argue, however, that the way the MC literature operationalizes affordances is still rather mechanical: A (perceived affordance(s)), in interaction with B (individual's goals), leads to C (individual's behavior and/or cognitions). Indeed, most empirical MC studies tend to study the (cor-)relations of certain (sets of) affordances with user's behaviors and/or cognitions. For example, comparing the degree of social presence in media differing in synchronicity (Tang & Hew, 2020), contrasting the degree of politeness in media channels differing on identifiability and networked information access (Halpern & Gibbs, 2013), comparing platforms that differ on visibility and persistence on the degree of jealousy they engender (Utz et al., 2015), or assessing perceived effectiveness of social support and relational closeness across platforms differing in the availability of "likes" or "upvotes" (Hayes et al., 2016). This affordances-based approach is also still individualistic. Although (almost) all the outcome behaviors and cognitions are interpersonal, that is, in reference to one or more other person(s) – expressing one's opinion to someone, sharing knowledge with someone, feeling jealousy towards someone, perceiving social support from someone – they all involve the assumption that the user and the other(s) they are interacting with should be considered as separate, relatively independent, entities ("me" and "you"). As a consequence of this approach, most empirical MC studies do not look at *inter*action: a co-operation between two or more individuals.

One Step Further: Structuration and Construction

We propose that it would be worthwhile to take Carr and Flanagin's approach a step further. We believe that people do not necessarily change behaviorally and psychologically only because the medium and its affordances channels their actions in a particular direction. At least a part of the behavioral and psychological change that occurs when using a medium is produced *collaboratively* with others. Inspired by the scientific traditions on the philosophy of mind and language (e.g., Searle,

1990, 1995) and sociological theory of knowledge (e.g., Giddens, 1984), we propose to look at mediated communication as joint action.

To illustrate the potential benefits of making this extra step, we can consider MC research that studies interaction in the sense of a back-and-forth action between two agentic interaction partners: research inspired by the hyperpersonal model of Walther (1996). This model shows how dating partners use the possibilities of mediated communication to present themselves in the most desirable way, how this behavior results in idealized impressions in the interaction partner, and how this partner subsequently confirms the idealized self-presentation in their behavioral reply. The resulting feedback loop substantiates and magnifies the idealized impressions that partners form from each other, resulting in hyperpersonal connections that feel extraordinary intimate. The model has received extensive support in many different contexts (Walther & Whitty, 2020). This model clearly captures the collaborative nature of mediated communication - the interaction behavior of one partner will be met with a (cognitive and behavioral) reply by the other, which will elicit another contribution of the first partner, etcetera. But the approach also assumes that such interactions are between interdependent actors with independent cognitions: people act in a certain way to achieve a personal social goal (to make a favorable impression), and they use properties of technology creatively to achieve their goal. The hyperpersonal actors in this model have desires and goals which shape their actions, and they form impressions of the other. Together this can produce romance. But they are not yet actors who have thoughts or feelings about who they are and what they are doing at a *collective* level.

Of course, such back-and-forth transactions play an important role in many different social situations (see also Fiske, 1992). In the context of extended social interaction, however, such ways of acting and thinking could lead to, in the words of Searle (2010), “infinite regress”: oftentimes it is impractical or even impossible for interaction partners to think about what they think that the other thinks that they think that the other thinks, and so on, ad infinitum. It is both more efficient and more productive for interaction partners to form cognitions about the actions that they are performing together. Searle calls this “collective intentionality”. Collective intentionality is not a combination of individual intentionalities but irreducibly collective (Searle, 1990, 1995). Thinking at the level of “us” is not only less effortful but also essential to effective interaction since interaction is joint action (Bavelas et al., 2000; Clark, 1996; Sebanz et al., 2006). Collective intentionality enables interaction partners to successfully coordinate their actions and achieve their collective goals.

This sense of where they are going and what they are doing is for an important part emergent from the social structure that people find themselves in. Social structure is a rather abstract term, originating from sociology, that broadly refers to the regularities of interactions within a given social entity (Calhoun, 2002). According to Structuration Theory (Giddens, 1984), social structure steers interaction but is also emergent from interaction as it needs to be negotiated and established by interaction partners together. First, social structure steers interaction by providing the context of rules and constraints in which individuals act. In other words, the social structure produces the norms that prescribe how the participants in the interaction ought to, or tend to, behave. Second, social structure is emergent from interaction as its content depends on interaction partners' behavior: when people act in line with the constraints the structure places on their behavior, the structure is reinforced, but when they act outside of the structure's norms, their behavior can be corrected, or the structure can be modified (see also Koudenburg et al., 2017, 2021). This process whereby people jointly establish norms has been referred to as social construction. So, any interaction behavior simultaneously defines and reflects the group norm, and interaction partners collectively shape and reshape the social structure in which they act.

At the start of the Digital Revolution, DeSanctis and Poole (1994) introduced Adaptive Structuration Theory (AST) to recognize that technology also "structures" human action. That is, technology brings with it a new set of rules and resources, or constraints and possibilities, for behavior. DeSanctis and Poole specified that this was non-deterministic as people can still use technology in different ways and thereby affect the structure that technology provides. This means that the use of technology can be an invention of the users themselves. There is consequently a continuous reciprocal influence of social and technological context on the structuration of technology.

The structuration theories remain somewhat opaque about the exact process of structuration and how this works within interactions. Research in the tradition of the Social Identity model of Deindividuation Effects (or SIDE model) has given more insight into this by empirically showing how, in the context of online interactions, interaction partners start to conform to the prevalent communication style and content over time (Postmes et al., 2000). In this way, norms that prescribe a particular use of technology in a specific context are induced from the interaction behaviors in that context.

The social construction and structuration perspectives essentially suggest a dynamical systems approach to (mediated) interaction (Vallacher et al., 2002). Participants in interaction become so intertwined with each other and the interaction

environment that we must consider them as a system of interdependent elements. This system is dynamic and in motion because the interdependent elements are subject to continuous change. By interacting with (one element in) the system, the entire system is changed. Interaction partners' behaviors are determined by this system, but also determine it. Here interaction behavior is not considered an outcome on an individual-level (like the mechanical approach) but part of an infinite cyclical process with collective consequences.

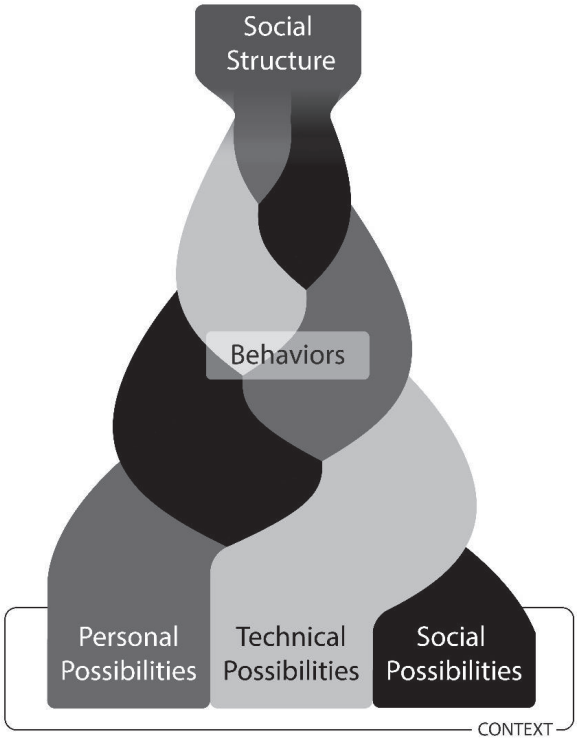
To take stock, the current approach to MC could be classified as individual or interpersonal and mechanical. Whereas this can teach us about the effects of the medium and/or its affordances, it might give an incomplete picture of the process of interaction by ignoring social dynamics. The alternative would be a collective and constructivist approach to MC. But, whereas this would allow us to learn about the interaction process, we cannot conclude much about the social effects of the usage of particular media. Since behavior is assumed to be "radically situated" (Barker, 1968, 1987), that is, embedded in a specific (medium) system or structure (Gibson, 1966), it is inherently incomparable with other (media) systems or structures. We propose that in-between these two perspectives lays fertile terrain to explore: where technology guides behavior(al dynamics), and behavioral dynamics shape the construction of a social structure. In this article, we seek to integrate the mechanical with the constructivist perspectives and propose that this enables us to learn about differences between media *and* the dynamical process within a specific medium. In fact, by studying the processes within media, we can explain the differences between media. The key to successful integration, we propose, is behavior.

SoDA: Theoretical Framework

The Social Dynamics Approach to Mediated Communication (SoDA) consists of a theoretical framework and a methodological framework. We will first describe the theoretical framework and outline the methodological framework in the next section. In short, the theoretical framework posits that the way participants behave, and thereby the form that their interaction takes, functions as a constant gauge of their understandings of their social structure. When behavior is in line with what is considered normative and expected based on these understandings, this reaffirms them. When behavior is *not* in line with social norms and expectations, participants conclude that there is a misunderstanding between them about who they are and what they are doing (Koudenburg et al., 2014; 2017). We assume that behavior is shaped by *social* possibilities and constraints (as proposed by the constructivists), but also (in line with the mechanical approach) that it is influenced by technological and personal possibilities and constraints. As can be seen in Figure 1 below, the theoretical framework posits that these three factors (technical, personal, and social;

the three strands of the braid) together shape interaction behaviors and thereby give meaning to the social structure as perceived by interaction partners (the tip of the braid). The interplay between technical, personal, and social constraints and possibilities takes place within and is influenced by a certain (social and medium) context (bottom square). Partner's assumptions about their social structure in turn affect their social norms, personal goals, and the way they make use of technology. Problems arise when technological possibilities and constraints limit interaction partners' ability to show behavior that is socially expected and accepted since this will destabilize the existing social structure and force it to change. We will now outline and explain each of the features of the framework in more detail.

Figure 1.
Schematic representation of the theoretical framework of the Social Dynamics Approach to Mediated Communication (SoDA).



To have an effective interaction, participants need to collaborate and therefore coordinate their actions (Bavelas et al., 2000; Sebanz, et al., 2006). In psycholinguistics, there is substantial theoretical and empirical work on how interaction partners either unconsciously align (Pickering & Garrod, 2004) or consciously coordinate (Clark & Brennan, 1991) their understanding of the meaning of their communication. Interaction partners need to have similar representations to understand what the other is talking about. But also more generally, interaction partners need to establish common understandings of the content and the process of their interaction. They need to align on *what* they are doing in the interaction: their personal and collective goals (cf. collective intentionality), and *how* they will do this: their mutual behavioral *expectations* or social norms (cf. social constructivism). This process of collaboratively shaping and updating shared understanding and shared viewpoints has been referred to as building common ground (Kashima et al., 2007).

Building common ground is not merely a process that facilitates the interaction. By collaboratively establishing what they do and how they act, interaction partners come to form a sense of who they are as a collective; what we call their perceived social structure. Acting in line with the mutual expectations about how they should behave to accomplish their collective goals, the co-acting interaction partners experience their interaction as well-coordinated and effortless and conclude from this that they share a social structure. This does not only rely on *what* is being said (i.e., interaction content) but also, sometimes even more so, on *how* things are being said: the form of interaction, and especially the subjective experience thereof (Koudenburg et al., 2014; 2017). The successful completion of even a simple short interaction sequence between individuals is diagnostic and informative about their perceived social structure. To use a metaphor: the body of interaction is experienced[1] by those individuals as the embodiment of the relationship that they have with each other (Kashima et al., 2007; Koudenburg et al., 2017).

There is a quite large literature that has demonstrated this to be the case. One strand of research has studied how people form a sense of social structure inductively, bottom-up, simply by having an interaction and forming a sense of shared understanding (e.g., Postmes et al., 2005; Swaab et al., 2007). For example,

[1] Is this always the case? Most certainly not. Humans' capacity for meta-awareness means that there are other ways of conversing (e.g., as a therapist, diplomat, etcetera) in which the pleasant and empathetic form of interaction is totally separate from the higher order cognitions about me, you and us. These people might be seen as manipulators with a double agenda, as they pretend to care about the social structure but in reality do not. And of course there might be other reasons why a person in the interaction is not psychologically involved in this dynamic: they might be preoccupied or distracted by more pressing matters, they might be on drugs, and so on. So it is important to clarify that what is written here applies to conversations between people who attend to the topic of the interaction and who are psychologically involved in the interaction: they must be *open* to the potential formation of relationships.

when some unacquainted participants in an online discussion discover that they agree about government funding for education, they can infer a sense of we-ness from this. But this process also works the other way around; deductively or top-down: by shaping the goals and expectations of interaction partners, the pre-existing perceived social structure affects the form of the conversation and thereby partners' sense of shared understanding (Koudenburg et al., 2017; Postmes et al., 2005; Swaab et al., 2007). For example, one may expect that among young males sexist talk is ok, and make a sexist statement in one's male friends group. When such a statement is rejected, one learns that this behavior is, in fact, not in line with the norms of this social structure, and the non-sexist norms are maintained. When the behavior is not rejected, sexism might become more normative in the group (Koudenburg et al., 2020). Thus, the perceived social structure affects the goals and expectations that partners bring to the interaction and thereby affects their behavior, but partners' behavior, and the extent to which this is in line with their goals and expectations, also affects the perceived social structure.

When interaction, even slightly, diverges from the way it is supposed to go, such as an unexpectedly long silence or people speaking simultaneously, interaction partners notice this immediately (see also expectancy violations theory, Burgoon, 1993; Burgoon & Hale, 1988). As the form of their interaction is so strongly tied to interaction partners' sense of social structure, a disturbance in this collaboration is experienced as a social threat (Koudenburg et al., 2017). Such disruptions raise questions about who they are and what they are doing as a collective. But note that what counts as a disruption depends crucially on interaction partners' expectations that are based on the existing social structure. For example, where unacquainted interaction partners take a silence or other delay to their conversation as a signal of misunderstanding and maybe even conflict, partners in a secure relationship might take these same signals as a sign of mutual understanding and validation (Koudenburg et al., 2014). As another example, when a subordinate interrupts someone of a higher status, this might threaten their established social hierarchy, but when it is the other way around, this will confirm their perceived social structure (Koudenburg et al., 2013b). It is worth noting that the effect sizes in this literature are considerable: this suggests that people are very sensitive to these social signals and that even very miniscule deviations from social expectations can lead to quite drastic re-assessments of social structure (see, for example, Koudenburg et al., 2013a; Koudenburg et al., 2021).

People's ability to behave in line with mutual expectations and thereby maintain the expected form of conversation, depends not only on their correct understanding and motivation to do so, but is also restrained and enabled by the

technical features of the context in which the interaction takes place, for example, online. The form of their interaction is importantly shaped by the technical constraints and possibilities that partners perceive within the interaction context (which can also be partly socially construed as partners might point out or demonstrate behavioral opportunities to each other). This could be termed "affordances" but we deliberately refrain from using this ambiguous term (e.g., Dings, 2021; Oliver, 2005) and opt for a more concrete description. We propose that conversation goals, perceived social norms and expectations, and perceived technological possibilities determine together what range of behavioral options people perceive they have within the conversation, which sets the boundaries for the kinds of behavior they are likely and willing to perform. That is, there is an interplay between personal (goals), social (norms and expectations), and technical (medium) constraints and possibilities for behavior in each interaction context. This implies that the three factors are not independent nor static but influence each other continuously within a conversation: when one of the factors changes, the other factors can change too. For example, when one would start to use Reddit, a platform for venting opinions, for dating and others would follow suit, this could drastically change what behavior is perceived possible and expected. Here personal goals influence the technological possibilities one perceives as well as the norms one associates with the platform. There are abundant real-life examples of this process of usage shaping technology too, such as the early online messaging service Minitel which was mainly intended for business communication but where erotically tinted messages ("messageries roses") became unexpectedly very popular (Feenberg, 1992).

Problems arise when interaction partners' goals and expectations are incompatible with the technical possibilities. For example, a person aiming to tell their far friend about their holidays over the phone, would expect to receive backchannel signals from this friend at certain intervals, like "hmhm" or "ok", as evidence of their continued attention. But a delay on the line prevents the friend from meeting this expectation, resulting in a disrupted conversation. Research shows that this will lead interaction partners to the conclusion there is a problem between them: there must be a misunderstanding between them and their perceived social structure is threatened, even though the external cause (a technical problem) is evident (Koudenburg et al., 2013a). So, collaborative behaviors are enabled and disabled by the interaction context but internally attributed, which has collective consequences as people think there is a relational problem between them.

In sum, social interaction is a collaborative act in which structures such as norms, conventions, and shared identities are formed, shaped, and maintained. There is not just what I want to say, or what you want to say, but there is a

continuous management of "us". As such, the interaction partners and their interaction environment form a dynamical system of interdependent elements, which is subject to constant change. But not everything is fluid and emergent. We propose that technological limitations can still impose restrictions on behavior and thereby influence the form and shape of the social interaction: technology is not neutral. What outcomes this ends up having, is a product of how the interaction unfolds, which also depends on interaction partners' personal goals and the perceived and practiced social norms. This integrated perspective means that making between-medium comparisons can be very informative, as long as we compare dynamic systems across different media and technological contexts.

SoDA: Methodological Framework

The theoretical framework outlined above implies a certain way of conducting research – a methodological framework. In this framework, we propose to study mediated communication by closely looking at actual conversational behavior and its collective causes and consequences within the dynamical system of an interaction. In line with the current MC literature, we put (micro-level) behavior center stage: how do people (strategically) act within the behavioral possibilities and limitations a certain interaction context provides? We additionally propose that a complete understanding of why these behaviors lead to certain outcomes is reached by analyzing the social dynamics that result from this behavior, and assessing the collective outcomes that follow from these social dynamics. Technical possibilities shape social dynamics by shaping how interaction partners act to realize goals and conform to expectations. The ensuing social dynamics shape the sense of social structure and therefore have collective outcomes.

Many current MC studies focus on isolating (sets of) affordances, what we call technical possibilities, to test their consequences for the interacting individual in experiments or observational studies, but we believe we need to understand how these technological possibilities are put to work in a specific social structure, made up of people that hold certain assumptions about who they are, how they relate to one another, and what they are doing as individuals and collectively. We need to study mediated communication in situ, because behavior and its meaning emerges organically in the ecology of the social setting. Indeed, the same technological possibilities might lead to differing, even contradictory, behaviors depending on the perceived social structure. Interaction partners can strategically use technical possibilities in different ways, enabling different forms of behavior, depending on their goals and the social expectations they (implicitly) associate with a particular interaction context. For example, being able to express oneself anonymously, creates particular opportunities for the expression of positive (e.g., prosocial, Shifrin &

Giguère, 2018) as well as negative (e.g., hostile and aggressive, Rösner & Krämer, 2016) emotions, depending on intentions or goals and the local social norm that prevails (see also Klein et al., 2007; Postmes et al., 1998). Also, interaction partners' interpretation of each other's behavior might radically differ depending on the interaction context in which it is perceived. Media environments have social norms, which may echo the social functions they fulfill. Whereas you may end your email to a colleague with, "best wishes, [your name]", writing this at the end of a WhatsApp message may be perceived as overly formal, and saying the same thing when you are meeting this colleague in person would be outright weird. Relational inferences of communication behaviors are thus partly medium and social context dependent.

This methodological framework invites us to compare how the same people with the same perceived social structure and the same goal(s) interact via different media (or contexts with differing technical possibilities), or, the other way around, to compare how people with differing social structures or personal goals interact within the same medium. This makes the close study of interaction partners' social behaviors and collective perceptions within media central, in such a way that cross-media comparisons can be made: how do interaction partners respond to the restrictions and possibilities of technology in their interactions and how is this behavior interpreted, attributed, and reacted to in the context of a certain social structure?

Operationalization

Our methodological framework proposes to measure social dynamics 1) by tracking behavior within individual interactions, at a micro-level, and 2) by assessing experiences at a collective level (such as conflict or love). Many MC studies rely on self-report surveys about media usage, big data, or other cross-sectional research (see, for example, Meier & Reinecke, 2021), but to analyze interpersonal dynamics, qualitative data are of central importance. We suggest to use mixed-method approaches and to complement big data and surveys with in-depth manual content or discourse analysis of interaction behaviors. Further, as we want to know people's behavioral intentions and interpretations of each other's behaviors within a certain medium context, we need, quantitative and qualitative, measures that can capture experiences of "us" in a single conversation.

First, the analysis of big data with complex and technically advanced methods has become increasingly popular in MC research. Big data often comprises countable metrics, such as number of likes or replies, or automated analysis of content (often based on word counts) as measures of interaction behavior (see, for example, a recent review of Reddit studies: Proferes et al., 2021). These data are readily available and easy to analyze, and, as there are so many datapoints, they can expose large-scale

patterns that cannot be identified otherwise and results might be highly reliable. The risk is, however, that they provide a somewhat simplified image of what happens in interaction. Even though the limitations of automated content analysis are widely recognized and many researchers stress the importance of always validating these with manual coding (e.g., Van Atteveldt et al., 2021; De Graaf & Van der Vossen, 2013), in reality, it appears that this micro-level checkup is often forgotten, especially in studies in computational sciences that work with big datasets (see also Song et al., 2020). Big-data analysis might be especially ill-suited for studying social dynamics within an interaction. For example, one might perform an automated analysis on the quality of relations on an online forum by counting the frequency of words classified as incivil behavior, such as "hate", "asshole", "f*ck". But, whereas this shows how negatively valenced the posts are, it does not tell us whether this negativity was aimed at other participants in the discussion or towards a common enemy, who might not be present in the conversation (e.g., the government). To be able to distinguish conflict from consensualisation, a more in-depth and context-aware analysis of interaction behaviors is needed, in which relations between posts take center stage. In sum, although big data can give important insights, we believe that even more valuable knowledge can be gained when we dare to let go of the big numbers and embrace the small and subjective for analyzing dynamics.

Second, there is a need for tools that can assess conversational experiences and collective perceptions (experiences of "us") within a single interaction. Currently, there are only a few such scales available, which is not surprising considering that the process of interaction and its social dynamics are rarely studied. Most scales used in the MC literature, and communication literature more broadly, measure general perceptions of and feelings towards either the self (e.g., self-esteem, Rosenberg, 1965; generalized communication apprehension, McCroskey, 1982), partner(s) (e.g., individualized trust, Wheeless & Grotz, 1977; partner responsiveness, Reis et al., 2018), or relationship (e.g., attraction, Montoya & Insko, 2008; relationship satisfaction, Hendrick, 1988). These scales are not specific to one interaction episode and do not allow a combined assessment of the self and the other and the relationship. For this reason, we have developed some alternative measurement instruments that assess people's appreciation of the social interaction with others. One example is the feeling heard scale that measures the experience of being heard within an interaction, an important component of which is the experience of mutual understanding with one's interlocutor (Roos et al., 2022b). Another example is Koudenburg et al.'s (2013a) measure of shared cognition which essentially assesses whether partners experienced a sense of common ground during their interaction (an example item is "I had the feeling my partner and I were on the same wavelength").

Beyond these quantitative measures of conversational experiences at a collective level, there is also a need for qualitative assessments because people's behavioral interpretations and intentions can differ greatly between different media contexts, in ways we cannot predict. This might take the form of post-conversation interviews (or open-ended survey questions) in which interaction partners are asked about their experiences of their joint action and their relationship. Moreover, since we are looking at social dynamics, it would also be very valuable to measure conversational experiences *during* interactions. There are tools for this that enable the coding of "the stream" of experiences in real-time (e.g., the mouse paradigm, Vallacher et al., 1994; see Jans et al., 2019 for an application in the context of a collective experience: sense of belonging).

The SoDA Applied: New Insight into Online Discussions

We believe that the Social Dynamics Approach to Mediated Communication outlined above can provide important new insights into existing MC research and theorizing. In this section, we will illustrate this with our recent line of research in which we applied the SoDA and that offers a new perspective on three prominent MC literatures: 1) media richness, 2) online disinhibition, and 3) online polarization. We will first summarize our research line and then outline how the resulting insights complement the three MC research areas.

Online Social Regulation

In this line of research, we attempted to uncover processes behind the apparent greater propensity of text-based online discussions to misunderstanding and polarization compared to FtF discussions (e.g., Anderson et al., 2018; Coe et al., 2014; Davis, 2009; Yarchi et al., 2021). We set out to find explanations for this between-medium difference by looking at interaction dynamics within media. We therefore closely examined text-based online and FtF discussions on controversial issues. In line with Social Information Processing theory (Walther, 1992), we assumed that most people seek to prevent conflict and maintain harmonious social relationships, also in online discussions (see also Papacharissi, 2004, who shows online incivility is rare). This is especially pertinent in the context of controversy because disagreement can threaten relationships (Brown & Levinson, 1987; Pomerantz, 1984).

In the interactions we studied, the *perceived social structure* consists of a group of strangers who likely expected to have a civil and constructive discussion about a controversial topic, aiming for consensus or common understanding. Participants' *personal goals* in this context are likely sharing their opinion and maybe even convincing others of that opinion. These personal goals are also

influenced by the perceived social structure, however, such that participants need to balance the aim of maintaining harmony with the need to share their opinion. In order to protect this social structure while accomplishing their personal goals, participants will have to show diplomatic behaviors: diplomacy is the *social norm*. We expected, however, that the *technological* constraints and possibilities of a text-based communication medium might not accommodate this norm. That is, communicative behaviors that serve a diplomatic function in FtF discussions might be more difficult to enact in a text-based medium. We expected this to have social consequences because non-normative behavior will lead people to question their understanding of the social structure: "Apparently, we are not all motivated to preserve social harmony and reach consensus". This might explain why it is more difficult to harmoniously navigate disagreements online. We focused on two behaviors that, according to the pragmatics literature, serve an important diplomatic function in FtF discussions, and that might be limited by the text-based medium: responsiveness and ambiguity.

Diplomatic Behavior FtF: Responsiveness

A first behavior with an important diplomatic function in FtF discussion is responsiveness. We defined this as the degree to which participants provide each other with instant feedback during interaction (Roos et al., 2020a, 2020b). People are responsive by sending backchannel signals during their interlocutor's speaking turn, for example, by interjecting words ("yes"), vocalizations ("hmm"), glances, or head nods (Beňuš et al., 2011; Clark & Brennan, 1991). By doing so the listener communicates that they are paying attention and encourage the speaker to continue. People also enact responsiveness by initiating a relevant next turn which contains an explicit or more implicit reaction or reference to the content of the preceding turn, for example, starting with a connecting word or utterance like "Yes, but…" or "And I would add…" (Clark & Brennan, 1991). Although these signals are often quite subtle, research suggests that interaction partners are very sensitive to these kinds of cues as ways of establishing and monitoring linguistic understanding of what is being said (Clark & Krych, 2004). But responsiveness also has a wider significance by conveying relational information: it shows interpersonal interest and is taken as signal of attraction and social harmony (Davis & Perkowitz, 1979). Moreover, recent research has shown that responsive and smooth interaction leads people to conclude that they agree on the discussed topic and everything between them is OK (Koudenburg et al., 2013b, 2017).

Diplomatic Behavior FtF: Ambiguity

Another behavior enabling relational maintenance amidst a FtF controversy is ambiguity. In such a "face-threatening" situation, people engage in indirect and

nonstraightforward communication (Bavelas et al., 1990; Brown & Levinson, 1987). Indeed, research shows that people in FtF discussions pre-empt conflict by ambiguating their message rather than expressing their disagreement clearly (Pomerantz, 1984). People can ambiguate with disclaimers (e.g., "I do not know for sure"), hedges (e.g., "maybe," "sort of"), and vocalizations that express doubt (e.g., a drawn out "hmmm") or tentativeness (e.g., "uhm," Brennan & Clark, 1996; Reid et al., 2003). Many of these expressions result from people probing what is acceptable to say while speaking. People do so by continuously gauging their partners' (mostly non-verbal) expressions of (dis)approval and editing their message accordingly. This tentativeness communicates tact and concern for the feelings of interaction partners (Roos et al., 2020a, 2020b). It also promotes constructive discussion and even consensualization by showing hesitancy and doubt about one's opinion – ambiguity implies ambivalence – and leaving room for interpretation of the message (Eisenberg, 1984; Pillaud et al., 2013). Essentially, ambiguity communicates that priority is given to relational maintenance over clearly and convincingly conveying an opinion.

Responsiveness and ambiguity thus enable FtF interaction partners to maintain a pleasant conversation and a good relationship amidst controversy. By enacting these diplomatic behaviors, participants conform to the social norms prescribed by the perceived social structure and thereby signal their endorsement of this social structure. That is, probably unknowingly, by making reference to their partners' contributions to the conversation (responsiveness) or by putting effort in phrasing a statement that will receive their partners' approval or at least will not offend them (ambiguity), participants jointly reinforce the perceived social structure in which they have a constructive discussion. We reasoned that people engaged in text-based online interactions most likely also subscribe to this social structure but the technological constraints might limit their ability to act accordingly.

Technology Affecting Diplomacy Online

In most text-based online chats, people cannot see each other and there is a relative lack of synchronicity. This does not provide interaction partners the opportunity to respond instantaneously to each other's comments. Rather, online participants may "talk" at the same time in instant chats (when they type and send messages simultaneously), or write a message after a long break in non-instant chats (e.g., on a discussion forum). These characteristics of the medium make it less likely that interaction partners will directly respond to each other's messages, or respond at all. An online "interaction" might thence end up looking more like a list of self-contained statements.

Further, due to the lack of visibility and synchronicity, most text-based online chats do not enable people to track each other's signals of (dis)approval and edit their message accordingly. Instead, the lacking visibility and synchronicity enable people to edit and complete their message *before* sending it to their interaction partners. Trying to be diplomatic in this context of technical constraints and possibilities, people likely prefer to think through and carefully formulate their messages before showing any of it to, and potentially offending, their interaction partners (see also Roos et al., 2022a). This results in precisely formulated and clear statements, which lack ambiguity.

As the behavioral change of increased ambiguity and increased responsiveness in controversial FtF conversations appears to be largely automatic, people might not be sufficiently aware of this behavioral change and its important function in maintaining harmonious relationships. People might therefore fail to acknowledge that their ability to act diplomatically is limited online and that this can have serious social consequences, and therefore might not seek to compensate for or counteract it. In fact, as will be described below, we even found that people actively resist interventions aimed at making their online communication more responsive and more ambiguous (Roos et al., 2022a). Thus, whereas participants might be motivated to have a constructive discussion online, they might not act accordingly due to the technological constraints and possibilities of the text-based medium. We tested this empirically by directly comparing text-based online and FtF conversations on controversial topics amongst small groups of unacquainted participants (Roos et al., 2020a, 2020b). An in-depth analysis of their conversation behaviors indeed showed that participants reacted less to each other's comments, i.e., were less responsive, and expressed themselves more clearly, i.e., less ambiguously, online compared to FtF, and this had social consequences.

Social Consequences of Limited Diplomacy

We found that participants in online discussions experienced their interaction as more disrupted, experienced more disagreement and more misunderstanding, and felt less solidarity towards their interaction partner(s) than in their FtF discussions (Roos et al., 2020a, 2020b). Importantly, participants' *perceptions* of disagreement were not related to the actual amount of disagreement that was expressed: they were caused mainly by the reduced responsiveness and reduced ambiguity in online discussions. Participants thus felt they disagreed not due to *what* their partners said but due to *how* they said it. Diplomacy is about how things are being said and thus about the *form* that participants' joint (inter)action takes. Technological constraints and possibilities affect this form, and it is this changed form that affects participants' experience of "us". This is in line with previous research showing that people infer

misunderstanding, dissent, and social rejection when the form of interaction, even slightly, diverges from their expectations, even if this is clearly due to factors beyond their partner's control, such as a delay in the communication channel (Koudenburg et al., 2013a, 2017).

We also found that participants felt ignored and thought that their partners were disinhibited online (Roos et al., 2020a). Participants felt their partners were more concerned with venting their own opinion than with listening to them. These perceptions were positively correlated with the reduced responsiveness and reduced ambiguity online. Thus, rather than acknowledging their partners' limited ability to express themselves tactfully, people tend to misattribute the lack of diplomatic behaviors to a lack of diplomatic motivation on the side of their interaction partners. Because partners are less responsive to each other's messages, they appear to be ignoring and disinterested in each other. Because partners formulate their opinion clearly, they may seem very convinced of the truthfulness of that opinion and not open to another opinion. Clearly communicating a disagreement could even be considered a sign of not shying away from conflict (see also Koudenburg, 2018). Indeed, feeling ignored and perceiving disinhibition were positively related to experiencing disagreement and a lack of solidarity (Roos et al., 2020a). That is, by being unresponsive and unambiguous, participants do not appear motivated to contribute to a constructive discussion climate where everyone can have their say and be heard and everyone's feelings are taken into account.

In sum, in the context of a controversial discussion where the *social norm* and therefore the (implicit) expectation is to behave diplomatically, and the *goal* is to maintain good relations while expressing one's opinion, the *technological* constraints of text-based media can limit people's ability to act in line with this norm and goal. When their partners' behavior is not in line with their norm-based expectations, people conclude that these partners do not endorse the same *social structure*; one in which they discuss opinions in a constructive and harmonious way.

Interaction Technology and Norms

This does not mean that social norms are static and overrule technology. In a series of studies, we found that people resisted interventions that aimed to make their online messages more responsive and more ambiguous (Roos et al., 2022a). For instance, with one intervention, we intended to give interaction partners insight into each other's thought process, and thus also the ambiguity within that process, by letting them converse in a chat environment that showed what they were typing in real time. We found, however, that participants actively worked against this intervention by formulating themselves as precisely as possible. Their interaction partners subsequently misinterpreted the resulting clear and self-contained

comments as a sign of strong conviction. In another study, we tried to make online interaction more ambiguous and responsive by transcribing a real FtF discussion as if it was a text-based chat and showing this to a sample of uninformed observers (Roos et al., 2021). We found that these observers experienced the FtF ambiguity in the online chats as weird and even as indicative of conflict. Together, these two papers suggest that, over time and with repeated interaction experiences, technological constraints and possibilities might affect the behavioral norms and expectations associated with that specific medium. This is a telling example of the mutual influence of the technological and social factors in the SoDA: technology use shapes norms and these norms subsequently shape the possibilities participants perceive the technology to offer, even when these possibilities suddenly change. And when interaction behavior does not match these technology-informed norms, this can again be attributed to a social failure. This does not mean that the social norms resulting from these technological constraints and possibilities are equally functional in maintaining social relationships. Although participants perceive a lack of diplomatic behavior online as normative, it is still interpreted negatively, in the sense that it reduces their experience of harmonious social relations. It seems as if participants adjust their behavioral expectations to the medium, but attribute the source of these new norms to the underlying social structure, rather than the technological constraints.

Three Areas of Insight

By applying the SoDA we gained important insights that shed new light on existing findings and explained contradictory results that were gained from more mechanical approaches to MC phenomena that have dominated the literature for the past decades. Specifically, we discuss our novel insights in three influential MC literatures: media richness, online disinhibition, and online polarization.

Media Richness

By zooming in on the dynamics of social interactions, our research contradicted one of the key tenets of Media Richness Theory (MRT; Daft & Lengel, 1986). While MRT was developed in the organizational or work context it has also been applied to other contexts, such as relational communication (Tong & Walther, 2015) and education (Shepherd & Martz, 2016). MRT distinguishes communication media based on their ability to transmit communication cues and thereby information from sender to receiver. The central tenet of MRT is that richer media can forward more cues and thus more information than leaner media. For example, the phone is considered a leaner medium than FtF communication because the former cannot transmit non-verbal signals while the latter can. MRT further holds that a primary driver for selecting a medium of a certain richness is the degree of

equivocality of the message. A highly equivocal message is unclear and more difficult to understand correctly. Equivocality can thus be considered a synonym of ambiguity. To prevent misunderstanding by the receiver, a highly equivocal or ambiguous message needs to be sent through a rich medium: in a rich context, communicators can reduce the unclarity by increasing the number of cues and thus the amount of information contained in the message (e.g., winking while making a sarcastic statement). So, according to MRT, unclear messages are undesirable because they can lead to misunderstanding between sender and responder, and richer media enhance a message's clarity by transferring more cues (Runions et al., 2013).

In contrast, we found that in situations that are prone to conflict because of potential disagreements, unclear messages are *desirable* because they signal people's endorsement of a social structure of harmonious and constructive discussion. We further found that the text-based online medium, which is leaner than FtF, stimulates users to write clear and succinct messages because *less* cues can be transferred. And surprisingly, in our studies, the richer medium offers communicators the opportunity to *reduce* their message's clarity, and they use this opportunity liberally. This sounds incompatible with MRT, but these contradictions can be resolved by considering the perceived social structure and the associated collective goal of conversation and how this relates to the content and form of messages. This can be connected to the two main components of a message that communication theorists commonly distinguish: the content message, i.e., the actual content about the topic, and the relational message, i.e., the perception of the sender towards the receiver (Adler et al., 2006; Trenholm & Jensen, 2008).

A clearly formulated message is probably very desirable in a perceived social structure where the collective goal is accurate information transmission, such as when planning a meeting with colleagues (at the time MRT was developed, 1986, this resembled the typical setting in which online communication took place). However, in the perceived social structure we studied where people aimed for a constructive discussion of a controversial topic, a clear message can come across as insensitivity to other people's opinions. So, there is an inverse relationship between the degree of clarity of appearance versus intentions here: by ambiguating a message content-wise, one clarifies, and prevents misunderstandings about, the message's underlying *social intentions*. Online, we found, is very clear in content, and this leads to unclarity on the relational level: social intentions are misunderstood. In the context of a controversial conversation, people appear to make an implicit assumption that "in the end, we agree and understand each other", but the practice of online conversations does not maintain this illusion because online comments explicitly

expose disagreements rather than hiding them under a cloak of ambiguity. Online, people take care to construct a well-formulated clear statement to prevent misunderstanding, but they do not seem to be aware that this clarity blurs their good relational intentions.

Thus, to come back to MRT: the FtF medium indeed makes the relational intent behind the message clearer but does this by ambiguating the message's content and form, which, in the context of a discussion about controversial topics, can serve social functions. This exemplifies how our approach differs from and complements the MRT literature. Where MRT focusses on how a communication technology can make a message more or less ambiguous, we add that this ambiguity has a communicative function in its own right as it is interpreted against the backdrop of a certain social structure.

Online Disinhibition

A second influential assumption is that the online medium disinhibits people (Suler, 2004). It is proposed that in many online communication environments, due to the affordance of anonymity, people lose the social constraints that normally guide their behavior. As a result, people dare to say anything they want, not limited by their concerns about the consequences for impression management and social relations. This can have both benign or prosocial and toxic or antisocial consequences, depending on the context. In a romantic context this may lead to increased self-disclosure that helps build relationships (e.g., Joinson, 2001). In the context of controversy this can lead to unbridled aggression that breaks relationships, like calling names or using rude language (e.g., Lapidot-Lefler & Barak, 2012). While evidence for the online disinhibition effect is inconsistent (Clark-Gordon et al., 2019; Lea et al., 1992; Spears et al., 2001), the idea that being online disinhibits remains widespread in the MC literature (e.g., Casale et al., 2015; Voggeser et al., 2018).

Our close examination of interaction behavior and the way this is socially interpreted led to a more nuanced view on online disinhibition. Notably, we did not find evidence for actual disinhibited behavior in the content of the online discussions: we did not observe any instances of aggressive or hostile language, swearing, derogatory names, etcetera. If anything, participants seemed to carefully construct their message before sending it (see also Roos et al., 2022a). However, due to their clear and unresponsive form, messages appeared more extreme, and senders appeared less concerned with how the message might be received by the recipients. Behaving diplomatically essentially communicates that one is inhibiting oneself, so participants misattributed their partner's unresponsive and clear expressions online to a lack of inhibition. Perceived disinhibition, in turn, was positively related to perceived disagreement and conflict (Roos et al., 2020a). The disinhibited

interaction partner does not seem to be invested in maintaining the social structure of constructive and harmonious discussion. Interestingly, we found that while participants perceived their partners as more disinhibited online than FtF, they did not consider themselves more (or less) disinhibited (Roos et al., 2020a).

This means that disinhibition can be observed, and even have consequences for social relations, without being enacted or intended. The online medium's characteristics might give rise to perceptions of disinhibition not by changing people's psychology but by changing their behavior and subsequent social dynamics. We suggest that it is not visual anonymity that makes people more or less self-aware (psychological change) and thereby leads to increased self-disclosure or rude language use[2], but the technical constraints of a medium directly affect the way people *communicate* (behavioral change), which has social psychological consequences. People are more outspoken and less responsive in text, which signals to their partners that they do not aim for a constructive and harmonious discussion. This suggests a shift away from considering disinhibition an individual psychological state that causes individual behavior to treating it as a form of interpersonal misunderstanding that is emergent from the social dynamics within interaction.

Online Polarization

There is considerable concern about increasing polarization in Western societies. According to many scholars and pundits, this trend is caused by social media and its algorithms. Consequently, there is a lot of empirical and theoretical research on online polarization, and how to explain and reduce this. Polarization is mostly operationalized as homophilic interaction patterns, aggravating opinion differences, or pronounced inter-group hostility (affective polarization; Iyengar et al., 2012). Research tends to measure online polarization indirectly based on self-reported users' perceptions, or, increasingly, on "hard" metrics from social media data (e.g., Iandoli et al., 2021), such as retweeting, hashtag use, following behavior (e.g., Garimella & Weber, 2017), network analysis (e.g., Kaiser & Pushmann, 2017), etcetera. Research often focuses on what medium characteristics correlate with online polarization, for example, selective exposure, that is, a lack of exposure to cross-cutting opinions (e.g., Mutz, 2006; Spohr, 2017), exposure to uncivil online discussion (e.g., Hwang et al., 2014), dissemination of fake news and misinformation (e.g., Del Vicario et al., 2017; Törnberg, 2018; Zollo, 2019).

The predominant approach in the online polarization literature thus appears to be similar to that of the broader MC literature described above: individual and

[2] We observed no effects of visual anonymity on any of the behaviors or psychological outcomes in our studies (Roos et al., 2020b).

mechanical. The literature tends to look at how technology affects individuals' polarized (i.e., extreme) behaviors and cognitions, sometimes depending on their goals: what does the technology expose the individual to, what does the individual do with the technology (who or what do they interact with), and/or how do they feel as a consequence of using the technology (e.g., affective polarization, Iyengar et al., 2012)? Our approach adds to this broad and rich literature by looking very closely at how technology affects dialogue.

Essentially, we found that the text-based medium limits people's ability to act in line with the expectations of a constructive discussion and that interaction partners, and outside observers alike (Roos et al., 2021), perceive this as polarization (Roos et al., 2020a, 2020b). Regardless of the actual difference of opinion between interaction partners, people conclude they must be polarized because the form of their interaction does not match that of a constructive discussion. We did not only observe this lack of dialogue in the lab, but also, even more pronounced, in large-scale discussions on social media. In a manual discourse analysis of discussions on Reddit we found that there was little responsiveness between the separate posts (Roos et al., 2022c). The discussions largely consisted of strings of individual viewpoints without much dialogue. People "replied" but barely *responded* to each other. We also found this in discussion groups on Reddit (so-called "subreddits") that are known for civil debate. The unresponsiveness might be due to, besides its text-based nature, the forum-like structure of Reddit which decreases the synchronicity of the medium and less synchronous media tend to be better suited for conveyance (broadcasting) than for convergence (dialogue) (Media Synchronicity Theory, Dennis et al., 2008). Notwithstanding these technological explanations, their behavior can make users appear mostly concerned with bringing across their own opinion rather than hearing the opinions of others and having a constructive discussion about that. That is, similar to the participants in our lab studies, people reading these posts might perceive polarization because of a lack of dialogue. Whereas an excessive focus on (clearly) venting one's own opinion at the expense of listening to others is often considered a *symptom* of polarization, we propose that this same phenomenon can be the *cause* of polarization online by inducing perceptions of polarization, which might, in the long run, cause real polarization (Davis & Dunaway, 2016). Rather than being deliberate and maleficent, polarizing behavior might be unintentional or even result from benevolent intentions (when people try to formulate themselves as precisely as possible to prevent misunderstandings, see Roos et al., 2022a). In short, when the technology restricts diplomatic behavior, people feel there is polarization because dialogue is impaired. Thus, to come back to online polarization, rather than looking at how online media might invite purposeful incivil and polarizing behavior or how online media might induce polarized views or negative impressions of others,

we propose to look at how online media directly affect interaction behaviors and how these behaviors affect participants' inferences about their social structure.

Implications and Future Directions

In this paper we introduced a new way of thinking about and studying mediated communication (MC) phenomena that takes social dynamics into account: the Social Dynamics Approach to Mediated Communication (SoDA). In short, besides the constraints and possibilities for behavior the mediating technology offers and their interplay with the user's goals and needs, we recommend MC researchers to also consider the social dynamics of the interaction itself which are importantly affected by and affect interaction partners' sense of who they are and what they are doing collectively (i.e., their perceived social structure). What this all boils down to is that when people communicate they *always* communicate information about their relationship. This is not new and in fact widely recognized in communication science (cf. the common distinction between the content and relational component of a message; Adler et al., 2006; Trenholm & Jensen, 2008), but we add that the mediating technology, by limiting and enabling communication behaviors, affects *what* people communicate about their relationship. This proposed adjustment, seemingly subtle and simple, implies a significant change in conceptualizing about and operationalizing MC phenomena.

In this paper, we illustrated how this new theoretical and methodological framework can be applied to a particular kind of interaction context: that of controversial discussions amongst strangers in text-based chats. But, we believe, the SoDA can be applied to any situation in which people are involved in an interaction. For example, it would be interesting to apply the framework to romantic relationship formation. How do people act romantically via different media? Maybe some media are inherently unsuited for romantic advances. Or maybe people adapt and behaviors that are considered neutral in one medium come to have a romantic or sexual connotation in another. Another example of an interesting application context would be a work setting where status differences are prominent in many social structures and behavioral expectations differ depending on the individual's relative standing (dominant versus submissive behaviors). Again, some media might be better suited to perpetuate this perceived social structure than others. Or people might adapt their behavior to technological limitations and use different status markers in different media.

We think that by taking into account the entire set of interconnected parameters specified in SoDA, the MC literature can develop a more nuanced and deeper understanding of within-medium interaction behaviors and their consequences for between-medium differences. This may help resolve some well-

known contradictions in the literature. To use one example close to our own line of research, there are contradictory findings about the consequences of the exposure to dissenting or cross-cutting views. Some researchers purport that cross-cutting exposure holds great promise for promoting the ideal of a deliberative democracy where all citizens participate in decision-making via reasonable discussion because research shows that exposure to dissent makes people more aware of the counter arguments to their own opinion (e.g., Price et al., 2002) and can reduce group polarization (Strandberg et al., 2019). But, on the other hand, others hold that cross-cutting exposure can be met with defensiveness and even exacerbate the increasing polarization visible in many societies (e.g., Bail et al., 2018; Settle, 2018). We propose that such outcomes may both occur and that which of the two prevails ultimately depends on the social dynamics of the interaction and how this exchange is interpreted in light of people's perceived social structure. Depending on the setting and dynamics, an essentially similar exchange can either be experienced as an argument between (political) opponents, as a constructive discussion between like-minded group members, or as venting of a set of individuals (see e.g., Koudenburg & Kashima, 2021). In fact, recent research shows that making people with opposing opinions aware of nonpolitical similarities between them (such as common interests and demographics) – in effect changing their perceived social structure from one of opposing partisans to similar peers – makes them feel more close towards each other, which increases their open-mindedness and decreases their opinion polarization (Balietti et al., 2021). Therefore, understanding the social dynamics better seems to be key to explaining the seemingly contradictory effects.

As another example, the effect of anonymity on self-disclosure (as a benign disinhibition effect) is inconclusive (Clark-Gordon et al., 2019; Nguyen et al., 2012). Some studies find that people self-disclose more online than FtF (Joinson, 2001; Lapidot-Lefler & Barak, 2015), but others show that they do not (Hollenbaugh & Everett, 2013; Schiffrin et al., 2010). This contradiction could potentially be explained by taking into account the interplay between technological, personal, and social possibilities in the SoDA and their relation with the perceived social structure. For example, on Tinder, the general assumption of many users appears to be that it is a platform for people looking for casual sex (e.g., Bulman, 2016). If this is indeed seen as a common goal by users, then once they have a "match" and start to interact via the chat, it is unlikely that they expect or give much self-disclosure because they do not expect or intend to build a relationship. In some sense, within the boundaries of the common understanding of what Tinder is for, it is *counter-normative* to be intimate and involved. In this way users together shape and maintain a set of understandings and practices with attendant goals and norms which shapes how users perceive and make use of certain affordances. At the same time, however, the

technology is not neutral: the possibilities of the platform encourage users to create profiles without in-depth personal details that lend themselves to being evaluated based on a quick scan of outward features (i.e., swiping photographs).

Other online dating platforms, such as eHarmony, have very similar characteristics and could potentially be used in a similar way, with users not disclosing much about themselves. But users might behave quite differently than Tinder users, because of the differing norms associated with eHarmony. The latter purports to attract a set of people with the goal of finding a serious relationship (see https://www.eharmony.co.uk/about/). Here people will therefore mutually expect and perform self-disclosures as this is needed to find the perfect match: the interactions engaged in by users thereby perpetuate the perceived social structure. Norms thus shape the use of the platform but, the other way around, the norms might also partly originate in the technological possibilities of this platform, for example, eHarmony shows its users metrics of the interests they have in common with their potential dates.

In sum, we have suggested an alternative approach to studying MC effects: the Social Dynamics Approach to Mediated Communication (SoDA). The SoDA integrates characteristics of more mechanical (deterministic) approaches to studying MC effects with social constructivist (indeterministic) approaches. The SoDA seeks to combine the strengths of both by studying up close how the technical features of the communication medium influence the inherently collaborative act of interacting. This allows us to study and better understand interaction partners' behavior and experiences within a certain communication environment with inherent possibilities and limitations for the their exchange. By closely observing how interactive behaviors unfold *within* communication environments, and by designing these observations in such a way they are comparable across media, we believe we can further enhance our understanding of why and how behaviors and experiences differ *between* media. Taking into account both the material limitations imposed by media as well as the social processes by which users seek to achieve their personal and social objectives (sometimes to overcome these limitations or sometimes to use them strategically), we believe, is one way for our field to further grow and develop its understanding of how MC interactions are transforming society at a rapid pace in ways that will no doubt continue to arouse controversy, but which clearly fascinate us all.