A Logic for Social Influence through Communication

Zoé Christoff

Institute for Logic, Language and Computation, University of Amsterdam
zoe.christoff@gmail.com

Abstract. We propose a two dimensional "social network plausibility framework" to model doxastic influence through communication in a social network. To do so, we combine two approaches: on the one hand, a hybrid logic setting, to model the social network itself (who is related to whom), and on the other hand, dynamic epistemic logic, to model the distribution of beliefs among agents (who believes what) and belief changes induced by communication events (what is said to whom and how do the hearers revise their beliefs). Combining both, we show how to design some particular communication protocols in this new framework to represent some level of social doxastic influence, assuming that the communicating agents are sincere and trust each other.

1 Introduction

Agents involved in a social network typically interact with the agents they are related to. By exchanging information, they influence those related agents and are influenced by them. If we consider the example of online social networks, agents communicate mainly with their "friends" or "followers". The same seems to apply in other examples of social relationships such as being colleagues or family members: people are influenced by people they interact with and the structure of who can interact with whom specifies a social network structure. In other words, agents who communicate are typically related by some social relationship (and hence, are part of s social network structure) and agents in a social network typically communicate with the agents they are related to according to the structure of a social network. Therefore, it seems quite natural to try and combine explicitly a social network structure with communication events in a unique framework to model the effects of social influence.

We propose a logical setting in which different scenarios of "social influence via communication" can be modeled. We are interested in reasoning about the doxastic states of agents in a social network and focus in particular on the communication protocols that can express different levels of social doxastic influence. Formally we use the tools and techniques of dynamic epistemic logic[1], combining

[1] See for instance [1, 7].

the work of Baltag and Smets on communication protocols for belief merge [3,4] with the work of Seligman, Liu and Girard on modeling social influence and peer pressure effects [8–10].

First, we build on the hybrid logic framework of Liu, Seligman and Girard [9], in the "Facebook logic" style of [10]. This setting combines a static social network model with an influence operator to represent how belief states change in a community according to the following *peer pressure principle*: every agent tends to align her beliefs with the ones of her friends. It is assumed that there are two situations in which an agent is pressured into *changing* her belief state. The first one is *strong influence*, the situation of maximal pressure to align, where *all* of my friends believe that ϕ, leading me to *revise* my beliefs so that I also believe that ϕ (after the successful revision with ϕ). The second one is *weak influence*, defined as follows: whenever I believe that ϕ but none of my friends believes that ϕ and some of my friends even believe that $\neg\phi$, I (successfully) *contract* my belief in ϕ. Moreover, while I'm being influenced by my friends' doxastic states, my friends are being influenced by mine, so that everybody is influenced by their friends' opinions all the time and *at the same time*. An important simplifying feature of this framework is that agents are influenced *directly* by their friends' beliefs, which corresponds to assuming that friends have access to each other's mental states, that they are in a sense *transparent* to each other. Given this definition of influence, it can be shown that in some configurations all agents will keep switching their opinions with their friends forever, while the other configurations will always reach a stable state at some point, after a finite number of repetitions of the influence operator. The language of the framework allows to characterize the stable configurations and the ones which will stabilize. A sufficient (but non-necessary) condition for stability is that the beliefs of everybody in the community are identical: once all friends agree, nothing changes anymore, since there is no pressure to align anymore.

Second, we adopt the perspective of Baltag and Smets in [3,4], investigating how a group of agents has to *communicate* in order to reach a state in which all agents "agree" on all their plausibility ordering, i.e, a state in which they have completely merged their opinions, in a way which reflects the relative importance of each agent. For instance, when an agent publicly announces a sentence which she believes to be true, she may convince the others, i.e., she may influence them into revising their beliefs with the announced sentence. The central idea of the beliefs merge protocols is that agents will speak in turn, according to a given rank of expertise, and announce to all the other agents all of what they privately believe to be the case. If the hearers trust the speaker enough, they will be influenced into revising their beliefs with the announced sentences, i.e, they will come to agree on what has been announced so far. This process can continue until a stable state is reached, a state in which none of what any agent believes would change anything anymore, if it was announced to the others. The reachability of such a global agreement stable state depends on how much the agents *trust* each other (do they revise with the announced sentences and if yes, how exactly?), on their *sincerity* (do they announce only sentences that they actually believe to be true?), and on the *exhaustivity* of the communication process (do they announce all non redundant sentences which they believe?).[2]

Both the above mentioned logical frameworks are concerned with representing the belief revision induced by what the other agents believe and with the reachability of a certain type of stable states. Moreover, in both settings, once all agents of a group agree, nothing changes anymore. However, while the first setting assumes direct (i.e., without explicit communication) bilateral and synchronic influence between all related agents, the second one assumes unilateral and diachronic influence on all agents through sequential public communication events. Moreover, while only the first setting allows to model *agents* and the social network explicitly, only the second setting allows to model *beliefs* and *belief revision* explicitly in terms of their underlying plausibility structures. We aspire to design a framework which can incorporate both aspects.

Combining both approaches, we propose a unified general "social network plausibility framework" in which the agents, their plausibility orderings and their social relationships are modeled explicitly and on which different communication protocols can be defined to represent different types of belief change under peer pressure. In the next sections, we will introduce the logical tools needed to build this new framework. In section 2 we introduce the two dimensional social network plausibility static framework. In section 3 we add the dynamics to our framework and show how the notion of strong influence from [9] can be redefined as a particular case of communication protocols in our framework, assuming a certain degree of trust in between friends, exhaustivity of the communication and success of the revision process.

2 A two dimensional social network plausibility framework

Our formal setting consists of two dimensions: a doxastic dimension and a social network dimension. To model the beliefs of agents, we follow [3,4]: we include in our models (epistemically) possible worlds and a subjective plausibility ordering relative to each agent and we include in our language the modal operator B for an agent's (simple) belief, defined as truth in all the states that the agent considers to be the most plausible.[3] To model the social relationships of the agents we

[2] [3,4] investigate which communication protocols lead the agents to merge different doxastic attitudes, and how they can merge their entire plausibility orderings. In this paper, we restrict ourselves to the case of *belief* merge, since belief this is the unique attitude considered in the influence setting of [9].

[3] For the time being we will limit ourselves to considering only simple belief. In future work, we will also consider other doxastic attitudes from [2], namely conditional beliefs B^{ψ}: belief under the condition that ψ; safe belief (or "defeasible knowledge") $\Box$: belief stable under revision with any new *true* information; and strong (or "robust") belief Sb: belief stable under revision with *any* (true or false) new information which is not known to contradict it, and irrevocable knowledge K.

follow [9, 10]: we represent agents and their social relationship explicitly in our model. To express things about this social dimension in an indexical way, our language contains a modality F, quantifying over friends, reading "all of my friends", nominals (each nominal is true at exactly one agent) as rigid designators and the operator $@_n$, which switches the evaluation point to the unique one satisfying the nominal n.

Definition 1 (Syntax). *The social network plausibility static language is the following, where $p \in \Phi$ is an atomic proposition and $n \in N$ is an agent nominal:*

$$\phi := p \mid n \mid \neg\phi \mid \phi \wedge \phi \mid F\phi \mid @n\,\phi \mid B\phi$$

Our two dimensional models are simply the result of embedding one dimension into the other one: a social network plausibility model is a (finite, multi-agent, pointed) plausibility model in which a social network frame is associated to each possible state.

Definition 2 (Social network plausibility model). *A social network plausibility model is a tuple $\mathcal{M} = (S, \mathcal{A}, \leq_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$, such that:*

- *S is a (finite) set of possible states,*
- *$\mathcal{A}$ is a (finite) set of agents,*
- *$\leq_a \subseteq S \times S$ is a locally connected preorder, interpreted as the subjective plausibility relation of agent a, for each agent $a \in \mathcal{A}$,*
- *$s_0 \in S$ is a designated state, interpreted as the actual state,*
- *$\asymp_s \subseteq \mathcal{A} \times \mathcal{A}$ is an irreflexive and symmetric relation, interpreted as friendship, for each state $s \in S$,*
- *$\|\cdot\| : N \cup \Phi \rightarrow \mathcal{P}(S \times \mathcal{A})$ is a valuation, assigning a set $\|p\| \subseteq S \times \mathcal{A}$ to every element p of some given set Φ of "atomic sentences" and assigning a set $\|n\| = S \times \{a\}$ for some unique $a \in \mathcal{A}$ to every element n of some given set N of "nominals".*

We inherited from [9, 10] an indexical semantics where every formula is evaluated both at a state $w \in S$ and at an agent $a \in \mathcal{A}$. For instance, assuming that p means "I am blonde", then BFp means "I believe that all my friends are blonde" and FBp means "all of my friends believe that they are blonde".

Definition 3 (Semantic clauses). *Let $\mathcal{M} = (S, \mathcal{A}, \leq_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$, $a, b \in \mathcal{A}$, $w, v \in S$, $p \in \Phi$ and $n \in N$. We denote by $\underline{n}$ the unique agent at which the nominal n holds, by $s(a)$ the comparability class of state s relative to agent a: for $t \in S, t \in s(a)$ iff $s \leq_a t$ or $t \leq_a s$, and we use the abbreviation $best_a\,\phi$ to denote the most plausible ϕ-states according to a: $best_a\,\phi := \{s \in \|\phi\| : t \leq_a s \text{ for all } t \in \|\phi\|\}$.*

$\mathcal{M}, w, a \vDash p$ *iff* $\langle w, a\rangle \in \|p\|$

$\mathcal{M}, w, a \models n \textit{ iff } \langle w, a\rangle \in \|n\| \textit{ iff } a = \underline{n}$
$\mathcal{M}, w, a \models \neg\phi \textit{ iff } \mathcal{M}, w, a \not\models \phi$
$\mathcal{M}, w, a \models \phi \wedge \psi \textit{ iff } \mathcal{M}, w, a \models \phi \textit{ and } M, w, a \models \psi$
$\mathcal{M}, w, a \models F\phi \textit{ iff } \mathcal{M}, w, b \models \phi \textit{ for all } b \textit{ such that } a \asymp b$
$\mathcal{M}, w, a \models @n\,\phi \textit{ iff } \mathcal{M}, w, \underline{n} \models \phi$
$\mathcal{M}, w, a \models B\phi \textit{ iff } \mathcal{M}, v, a \models \phi \textit{ for all } v \in S \textit{ such that } v \in best_a w(a)$

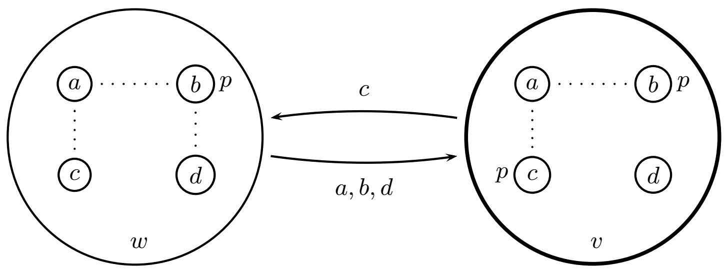

Fig. 1. A social network plausibility model with four agents a, b, c, d and two possible worlds w, v, where v is the actual world. The dotted lines represent the friendship relationhsip and the labeled arrows represent the subjective plausibility orderings of the agents (pointing towards worlds which are at least as plausible, loops are omitted). Agent c wrongly believes that agents b and d are friends.

3 Influence Dynamics

Let us now consider how models change under social influence through communication. We start by imposing some simplifying assumptions. First, only friends communicate.[4] What change is induced by communication between friends? In general, depending on how much an agent *trusts* the source of new information, she can transform her belief state in different ways.[5] We assume that friends trust their friends: whatever any of my friends announces, I revise my beliefs with it.[6] More precisely, we assume that friends trust their friends strongly enough to perform a *radical revision* with any formula ϕ announced. The operation on plausibility models corresponding to this strong level of trust is *radical upgrade* or "lexicographic upgrade" $\Uparrow \phi$ [2,5], which promotes *all* $\|\phi\|$-worlds so that they

[4] To simplify, we consider here only cases of public communication, since these are the cases considered in the protocols proposed in [3, 4]. This is only a starting point and we will relax this assumption and introduce a distinction between the insiders of some private communication, the friends of the announcer, and the outsiders (everybody else).

[5] See [5] for the definition of different types of upgrades corresponding to different levels of trust.

[6] We do not restrict our setting of influence to formulas for which revision is successful, unlike [9].

become more plausible than all $\|\neg\phi\|$-worlds (in each of the agent's information cell), while keeping everything else the same. We define the operation on social network plausibility models resulting from public communication in the obvious way: each agent revises its plausibility ordering (within each information cell) and everything else stays unchanged.

Definition 4 (Joint radical upgrade). *$\Uparrow \phi$ is a model transformer which takes as input $\mathcal{M}= (S, \mathcal{A}, \leq_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$ and outputs $\mathcal{M}'=(S, \mathcal{A}, \leq'_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$ such that:*

> $s \leq'_a t$ *iff either* $(s, t \notin \|\phi\|$ *and* $s \leq_a t)$ *or* $(s, t \in \|\phi\|$ *and* $s \leq_a t)$ *or* $(t \in s(a)$ *and* $s \notin \|\phi\|$ *and* $t \in \|\phi\|)$.

How do agents with such a level of trust have to communicate to reach a state where they all have the same beliefs? The assumption in [3, 4] that agents speak in turn, given some expertise ranking, allows to define the following *lexicographic belief merge protocol*: first, the agent with the highest rank announces that ϕ, for every (non-equivalent) ϕ that she believes. Then, the agent with the second highest rank does the same, and so on, until a state of global agreement is reached, i.e., a state such that nothing any agent could sincerely announce would change anything anymore.[7]

We adapt this protocol to accommodate the indexicality or our setting, i.e., we make sure that when an agent a announces ϕ, the hearers revise with $@_a\phi$ and not directly with ϕ. This seems to reflect in a natural way the indexicality of real-life communication. Note also that this protocol trivially requires that all agents are friends with each other.

Definition 5 (Beliefs lexicographic merge indexical protocol).

$$\rho_a := \prod\{\Uparrow @_a\phi : \|@_a\phi\| \subseteq S \times \mathcal{A} \textit{ such that } \mathcal{M}, w, a \models B\phi\}$$

$$\rho_b := \prod\{\Uparrow @_b\phi : \|@_b\phi\| \subseteq S \times \mathcal{A} \textit{ such that } \mathcal{M}_{[\rho_a]}, w, b \models B\phi\}$$

$$\textit{etc for all } c \in \mathcal{A}$$

where $\prod$ is a sequential composition operator and $\mathcal{M}_{[\rho_a]}$ is the new model after all agents have performed a radical upgrade with each formula announced by a.

In the reminder, we will show social influence as given in [9] can be reinterpreted in our framework as some particular communication protocols. Strong influence (I_s) is defined in [9] as the situation where *all of my friends* believe that ϕ: $I_s\phi := FB\phi$. This situation causes me to revise my beliefs with ϕ: assuming that revision is successful, whatever my initial state was, I will come to

[7] Here we consider a protocol to merge only beliefs, since belief is the only attitude modeled in [9] but [3, 4] actually consider how to merge the entire plausibility relations of all agents.

believe that ϕ and (assuming they did not change their mind in the meantime) *agree* with my friends.

Let us consider the simplest example of strong influence: some agent a is the other agents' only friend, has the highest expertise rank, and believes that ϕ. Translating this into a communication setting, a will announce that ϕ and the others will be (strongly) influenced into revising their beliefs with ϕ, *evaluated at agent* a. For instance, if a announces "I am blonde"(p), which is equivalent to her announcing "a is blonde" ($@_a p$), her friends come to believe that a is blonde ($@_a p$). They now agree on whatever was announced (they all believe $@_a p$). This is, therefore, at the same time, the simplest case of beliefs merge, and the simplest case of strong influence (as long as we only consider the case of a's friends and ignore what happens to agent a for now). This can be represented by a one step version of the beliefs lexicographic merge indexical protocol given above:

Definition 6 (One-to-the others unilateral strong influence protocol).

$$\rho_a := \prod\{\Uparrow @_a\phi : \|@_a\phi\| \subseteq S \times \mathcal{A} \textit{ such that } \mathcal{M}, w, a \models B\phi\}$$

where $\prod$ *is a sequential composition operator*

Assume now that all of agent a's friends are friends with each others. Agent a is in a state of strong influence with ϕ if and only if everybody else agrees on believing ϕ, in which case she will revise her belief with ϕ. But before she revises her beliefs, a first needs to be aware of what her friends believe. We define the following protocol, in which, unlike in the above, agents have to announce *that* they believe something, whenever they *did initially* believe it. This will result in a believing that all of her friends believe something if they actually did. We still need a to revise her beliefs with ϕ. So we add the last step of the protocol, fundamentally different from the rest of the protocol, representing the reasoning of agent a, the conclusion she reaches, announcing to herself (and thereby to her friends) that ϕ.

Definition 7 (The others-to-one unilateral strong influence protocol).

$$\rho_b := \prod\{\Uparrow @_b B\phi : \|@_b B\phi\| \subseteq S \times \mathcal{A} \textit{ such that } \mathcal{M}, w, b \models B\phi\}$$
$$\rho_c := \prod\{\Uparrow @_c B\phi : \|@_c B\phi\| \subseteq S \times \mathcal{A} \textit{ such that } \mathcal{M}, w, c \models B\phi\}$$
$$\textit{etc, for all } d \in \mathcal{A} \textit{ such that } \mathcal{M}, w, d \models \langle F\rangle a$$
$$\rho_a := \prod\{\Uparrow @_a\phi \textit{ iff } \mathcal{M}_{[\rho_b;\rho_c,\ldots]}, w, a \models BFB\phi\}$$

where $\prod$ *is a sequential composition operator and* $\mathcal{M}_{[\rho_b;\rho_c,\ldots]}$ *is the model resulting from the successive revisions (by all friends) with each of the formulas announced by each of them.*[8]

[8] The definition of strong influence has the counterintuitive consequence that if ϕ means "I am blonde" and if all of my friends believe that they (themselves) are

4 Conclusion and further research

So far we have considered only the case in which agents are under maximal and unilateral influence. We have ignored weaker cases where not all but only a significant proportion of friends believe something, and we have ignored the fact that *while* being influenced by my friends, I influence them too. This is a crucial difference between the way the communication protocols from [3, 4] are defined (agents speak in turn) and the way the social influence operator from [9] is defined "globally" and synchronically (all agents influence their friends while being influenced by them). To faithfully translate this global notion in the present communication setting, we would need to allow all friends to have the same expertise rank and therefore *speak at the same time* and we would need to define a corresponding "parallel" actions operator. We have also restricted ourselves to examples with *public* communication. In future work, we will consider private forms of communication (public announcement restricted to the group of the announcer's friends).

As mentioned in the introduction, in the setting of [9], friends are influenced directly by each other's *beliefs*, as if they had direct access to other agents' minds, as if they were "transparent" to each other. Our first goal here has been to show that this direct influence notion corresponds to a particular case of communication, where *sincerity, trust and exhaustivity replace transparency*, allowing agents to *get* access to each other's mental states. However, this assumption prevents such a framework from modeling some particular social phenomena where I am influenced not directly by what the others actually believe, but by what I believe that they believe, leaving some space for error or uncertainty. Our communication setting for influence should therefore be generalized beyond sincerity to properly distinguish between what an agent *privately believes* and what she *shares* with others (what she announces to others, what she expresses, what she seems to believe according to her behaviour, etc.[9]). In other words, in addition to *adding* communication to regain a transparent doxastic influence setting as we have started doing here, it would be interesting to focus on what agents can typically access (observe) of each other: their behaviour, whether it reflects their private beliefs or not.

Acknowledgments I would like to thank Johan van Benthem, Jens Ulrik Hansen, Emiliano Lorini, and Sonja Smets for their suggestions and comments during the elaboration of this short paper. An early version of this work has been presented at the Seventh Workshop in Decisions, Games & Logic and I would

blonde, I end up believing that I am blonde too. One solution is to restrict the above protocol to formulas ϕ whose truth do not depend on the evaluation agent, characterized by making the following valid: $@_a\phi \Leftrightarrow @_b\phi$.

[9] For a dynamic hybrid framework allowing such a discrepancy between what an agent believes and what she seems to believe, see [6].

like to thank the anonymous referees of both DGL and EUMAS/LAMAS for their valuable feedback.

The research leading to these results has received funding from the European Research Council under the European Communitys Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 283963.

References

1. A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. In *Proceedings of TARK'98 (Seventh Conference on Theoretical Aspects of Rationality and Knowledge)*, pages 43–56. Morgan Kaufmann Publishers, 1998.
2. A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory*, volume 3 of *Texts in Logic and Games*, pages 9–58. Amsterdam University Press, 2008.
3. A. Baltag and S. Smets. Protocols for belief merge: Reaching agreement via communication. volume 494 of *CEUR Workshop Proceedings*, pages 129–141, 2009.
4. A. Baltag and S. Smets. Protocols for belief merge: Reaching agreement via communication. *Logic Journal of the IGPL*, 21(3):468–487, 2013.
5. J. Benthem van. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 14:129–155, 2007.
6. Z. Christoff and J. U. Hansen. A two-tiered formalization of social influence. In D. Grossi, O. Roy, and H. Huang, editors, *Logic, Rationality, and Interaction*, volume 8196 of *Lecture Notes in Computer Science*, pages 68–81. Springer Berlin Heidelberg, 2013.
7. H. Ditmarsch van, B. Kooi, and W. Hoek van der. *Dynamic Epistemic Logic*. Springer, 2006.
8. Z. Liang and J. Seligman. A logical model of the dynamics of peer pressure. *Electronic Notes in Theoretical Computer Science*, 278(0):275 – 288, 2011.
9. F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*. Special Issue on Social Epistemology, C. Proietti and F. Zenker, editors, to appear.
10. J. Seligman, F. Liu, and P. Girard. Logic in the community. In M. Banerjee and A. Seth, editors, *Logic and Its Applications*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188. Springer Berlin Heidelberg, 2011.