Logical dynamics of belief change in the community

Jeremy Seligman Patrick Girard Fenrong Liu

Abatract: In this paper we explore the relationship between norms of belief revision that may be adopted by members of a community and the resulting dynamic properties of the distribution of beliefs across that community. We show that at a qualitative level many aspects of social belief change can be obtained from a very simplistic model, which we call 'threshold influence'. In particular, we focus on the question of what makes the beliefs of a community stable under various dynamical situations. Besides, we consider refinements and alternatives to the 'threshold' model. The most significant alternative is to move to consideration of plausibility judgements rather than mere beliefs. We show first that some such change is mandated by difficult problems with belief-based dynamics related to the need to decide on an order in which different beliefs are considered. Secondly, we show that the resulting plausibility-based account results in a dynamical system that is non-deterministic at the level of beliefs. Nonetheless, the plausibility-based account lacks certain intuitively desirable features, such as the preservation of the transitivity.

Key words: Belief revision, belief influence, community, plausibility judgement

When one moves from reasoning about what one believes or knows to what other people believe or know, certain conceptual distinctions become quickly necessary. Perhaps the most well-known of these is the distinction between common knowledge (everyone knows and knows that they know, etc.) and distributed knowledge (together, we would know). The multiplicity of knowers also encourages a shift of topic, from procedures for revising one's own beliefs to those for determining the opinion of a group. When one considers, in addition, social relationships, other topics related to the transmission of knowledge become relevant. This paper considers one of these in particular: the influence on one's beliefs of other agents to whom one is socially related.

Explicit modeling of social relationships plays a significant role in social psychology, artificial intelligence and economics. The seminal [6], which studies the patterns of interpersonal relations for groups in terms of social power, has led to many more recent mathematical and also computational models. For example, motivated by applications to marketing, the question of how to find an influential member of a network and how to maximize social influence have been recently studied in [5] and [9]. Social networks have been extensively studied also in economics (see, e.g., [8]) and there is a new field of 'social simulation' which uses computational models to predict and explain social dynamics (see, e.g., [7]). Yet little attention has been paid to the norms that guide people's reasoning about social relationships. Perhaps this is because of the individual-centered history of epistemology. Other people are inherently unreliable, and so one cannot have purely logical grounds for changing one's belief in response to the opinions of one's peers. Nonetheless, the fact that people do change their beliefs in this way allows information to spread along social channels, and this is of epistemological significance and benefit (perhaps) to the community. Moreover, an understanding of how we can reason about this spread is of interest. Here we will build on [12] and especially [14] in exploring the consequences for logic of some simple assumptions about social belief change and propagation.

In Section 1, we introduce an account of how standard models of belief revision, in the tradition of [1], can be extended to models of social influence on one's beliefs. In particular, in Section 1.2, we introduce a specific model based on some fairly conservative assumptions about the thresholds required for us to change our beliefs when influenced by others. This is used to demonstrate a variety of dynamic phenomena, that can be analysed logically using the method of automata (from [14]). In Section 2, we examine more closely the question of what makes a community's beliefs stable with respect to social influence, including consideration of de-stabilizing changes such as when an individual changes her beliefs unilaterally (Section 2.1) and when new social relationships are formed or dissolved (Section 2.3). We also consider the effect of these changes on aggregations of belief across the community (Section 2.2). In Section 3, we consider various alternatives to the simple model of threshold influence, in which those with whom we are socially connected are ranked in some way as more or less reliable. Finally, in Section 4, we note some problems with the reliance on the single proposition revision/contraction model we have inherited. An alternative based on plausibility relations (from [4] and [13]) is explored. Although the procedure is completely determined by the distribution of plausibility judgements across the community, if one looks only at the distribution of *beliefs*, it is non-deterministic, in an interesting way. We also note, in a manner similar to [14], the conflict between these mechanisms for social influence based on plausibility and the requirement that plausibility is transitive.

1 Doxastic influence

To be influenced by my friends is to change my beliefs so that they correspond better to theirs. To begin with, we will consider influence regarding a single proposition p. If I do not believe p and some significant number or proportion of my friends do believe it, there are several ways I could respond. I could, of course, ignore their opinions and remain doxastically unperturbed. But if I am influenced to change my beliefs there are at least two ways of doing so: I may *revise* so that I too believe p or (more cautiously) merely *contract*, removing my belief in its negation $\neg p$. We will write Rp for the action of revision and Cp for the action of contraction.[1] The only assumptions we will make about revision and contraction is that they are 'successful.' This means that after I perform the action Rp, I will believe that p, and after I perform the action Cp, I will not disbelieve p (i.e., I will not believe $\neg p$).[2] In logical terms, this means accepting the following as axioms:[3]

$$[R\varphi]B\varphi$$
$$[C\varphi]\neg B\neg\varphi$$

Now whether or not I change my beliefs in response to my friend's opinions and if I change them, whether I revise or merely contract depends on at least two things: 1) my own attitude regarding p, and 2) the cohesiveness of my friends' beliefs concerning p. The more cohesive an opposition I face, the more pressure I have to change. But also, if I start with an open mind about p, I may be more easily influenced than if I hold a strong contrary opinion. The particular balance of these factors varies from person to person and even from belief to belief. It may also be partly determined by higher-level beliefs about the reliability of one's friends or different matters. So instead of committing ourselves to a particular theory of influence, we will merely draw a distinction between two kinds of influence: that which leads, respectively, to revision and to contraction. In the case that I am influenced to revise my beliefs positively in favour of p, we will say that I am *strongly influenced to believe* p, and write this as $\mathsf{S}p$. There may be other reasons to revise my beliefs or to keep them the same, but with regard to social influence alone, *ceteris paribus*, the condition $\mathsf{S}p$ is necessary and sufficient for me to revise. Likewise, when I am influenced merely to contract my belief in $\neg p$ (if I had one), without necessarily coming to believe p, we will say that I am *weakly influenced to believe* p, and write this as $\mathsf{W}p$. We will also refer to the corresponding negative conditions of being strongly or weakly influenced to believe $\neg p$, written as $\mathsf{S}\neg p$ and $\mathsf{W}\neg p$, respectively; and we will assume that it is not possible to be simultaneously (strongly or weakly) influenced to believe both p and to believe $\neg p$.

With this terminology and notation, we can define a general operation of social influence regarding p. My being *influenced* regarding p, written Ip, is for me to revise my beliefs so as to believe p when I am strongly influenced to do so; otherwise, to contract my belief in $\neg p$ (if I have one) when I am weakly

[1] Proponents of any one theory of belief change may read Rp and Cp according to their favorite theory. The AGM account of [1] is certainly good enough for our purposes, but nothing we say here will depend too much on the details.

[2] Although success is accepted as a postulate of many accounts of belief change, including AGM, it does impose some limitations. In particular, many higher-order propositions such as the Moore-like propositional form "p but I do not believe p" are problematic.

[3] The general framework here is dynamic logic, in which an expression of the form $[\pi]\varphi$ is a formula that means 'after performing action π, φ is the case'.

influenced, and similarly for $\neg p$. If I am not even weakly influenced, my beliefs will remain unchanged. More concisely, Ip can be defined as the program[4]

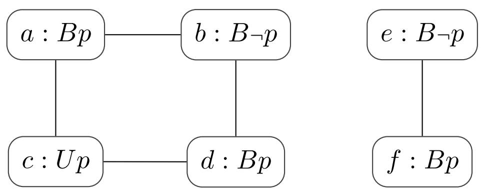
$$\texttt{if } \mathsf{S}p \texttt{ then } Rp \texttt{ else if } \mathsf{W}p \texttt{ then } C\neg p;$$
$$\texttt{if } \mathsf{S}\neg p \texttt{ then } R\neg p \texttt{ else if } \mathsf{W}\neg p \texttt{ then } Cp$$

From a logical point of view, this means that a logical system for R, C, S and W can easily be extended to a system for I also, using a standard treatment of PDL (propositional dynamic logic). Although we will not be exploring technical logical issues in this paper, this property will allow us to provide a reduction of statements about the dynamic properties of social belief change in the underlying modal language, as a result of the known reduction of iteration-free PDL expressions.[5] In the sequel, we will be offering various specific accounts of strong and weak influence, which enable a further reduction to a language that only contains operators for belief and the social relationships that structure our communities. Readers unconcerned with such technical niceties can cheerful ignore this talk of reduction, which will not play a role in what follows.

1.1 The dynamics of influence

There are three possible doxastic states of an agent with respect to the proposition p: belief (Bp), disbelief ($B\neg p$) and no belief ($\neg Bp \wedge \neg B\neg p$), which we abbreviate as Up. To discuss the distribution of these states among friends, we will use the framework of 'logic in the community,' introduced in [12], in which friendship is taken to be a symmetric and irreflexive relation. That is to say, I am a friend of any friend of mine (symmetry), and I am not one of my own friends (irreflexivity). We do not assume that friendship is transitive, so it is quite possible that my friends have friends who are not my friends. A set of agents related by friendship will be called a *social network*. A subset of agents that are connected by friendship, in the sense that for any two agents, there is a chain of friends that connect them, is said to be a *community*.

The distribution of doxastic states within a network can therefore be depicted by diagrams such as the following:

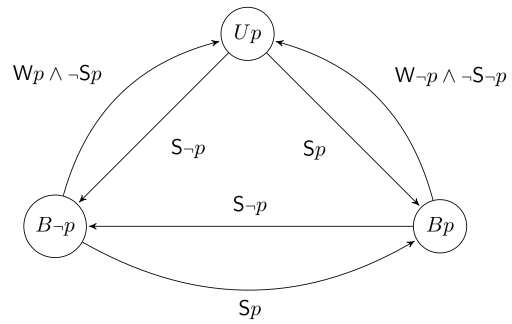


[4]The order of the positive and negative clauses is unimportant under our assumption that one cannot be both influenced to believe p and influenced to believe $\neg p$.

[5]This is an obvious consequence of standard axiomatic presentations of PDL, such as Definition 4.78 in [3].

This represents a network of six agents, clustered into two communities. Agent a believes p and has friends b and c; agent b disbelieves p and has friends a and d; and so on. As well as describing the doxastic states of agents (as Bp, $B\neg p$, or Up), we will describe their position in the social network using the symbol F to mean 'all my friends'. Thus FBp means that all my friends believe p, which in the example above is a true description of agents b and c but not of agents a and d. The dual operator $\langle F\rangle$ means 'some of my friends', so that, for example, $\langle F\rangle Up$ means that at least one of my friends is undecided about p. This is true of agents a and d but not of b and c.[6]

Now, given the success axioms for revision and contraction, each agent's doxastic state will change in a deterministic way under those operations. Revision with $\neg p$ will take her from state Bp to state $B\neg p$; mere contraction, from Bp to Up, and so on. Moreover, if we assume that the triggering conditions of strong and weak influence depend only on the distribution of doxastic states among agents in the network, this distribution will change under operation Ip in an entirely deterministic and local fashion. Careful analysis of the definitions given above shows that the dynamics of influence is characterised (for each agent, locally) by the following finite state automaton:[7]



The states of the automaton are the possible doxastic states of the agent. The transitions are labelled by mutually exclusive influence conditions. For example, if an agent believes that p, represented by the state labelled Bp, and is weakly (but not strongly) influenced by her friends to believe that $\neg p$, as represented by the label ($\mathsf{W}\neg p \wedge \neg\mathsf{S}\neg p$), then after she contracts her belief that p, she will be in state Up, undecided about p.

[6]The technical details of this language will not be relevant to our present purposes, so we will not go into them here, referring the reader to [12] for further details.

[7]Analysis of social logical dynamics by finite state automata was used in [14] to show that some interesting dynamic properties (such as the eventual convergence to a stable distribution) can be expressed in terms of operators similar to those we are considering here, but in the domain of preference rather than belief. Here we are providing a slightly more general characterisation for belief change, which does not depend on any particular account of revision, contraction, strong or weak influence. As in [14], it is important to realise that the machine is not the definition of a dynamical system but a tool to anaylse what is already implicit in the definition of the logical operators.

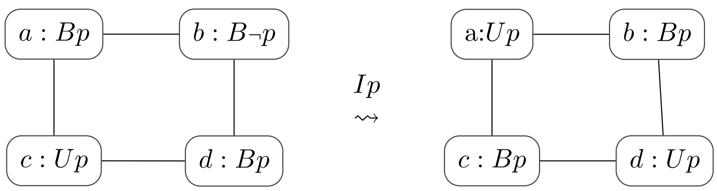
The reader may like to refer to this automaton to check the numerous examples that follow.

1.2 Threshold influence

To give examples of influence in action, we must provide an explicit account of strong and weak influence. A first guess is that strong influence requires cohesion among some threshold proportion of one's friends. Conservatively, we'll assume a threshold of 100%, meaning that I am strongly influenced to believe p iff all of my friends believe p (and at least one of my friends believes p). The parenthetical clause makes social hermits immune to strong influence, which is surely correct. Strong influence to believe $\neg p$ works similarly. For weak influence, we'll suppose that one must have at least one friend who believes and that the number of friends who disbelieve is no greater than some threshold. Again, conservatively, we'll assume that to be 0%. This means that if I am not strongly influenced to believe that p then I am weakly influenced iff none of my friends believe $\neg p$ (and at least one of my friends believes p).[8] This account of strong and weak influence is captured with the following axioms:

$$S\varphi \leftrightarrow (FB\varphi \wedge \langle F\rangle B\varphi)$$
$$W\varphi \leftrightarrow (F\neg B\neg\varphi \wedge \langle F\rangle B\varphi)$$

Example 1:



In this example, agents a and d both believe p and are weakly influenced to drop this belief, since all of their friends do not believe p (b disbelieves and c is undecided) and one of their friends, b, believes $\neg p$. Thus $W\neg p$ is true of a and d, and under the operation of social influence, Ip, they both contract their believe in p and become undecided. By contrast, agents b and c are strongly influenced to believe p by their friends, a and c. Thus Sp is true of b and c, and under the operation Ip, they both revise so as to believe p.

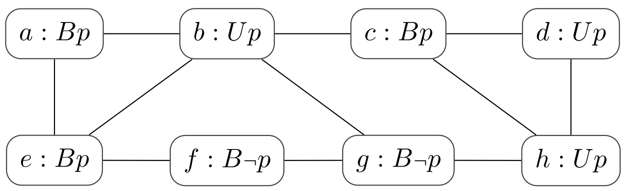
[8]This account of strong and weak influence is more-or-less parallel to that given for preference dynamics in [14].

2 Stability and flux

Example 1 shows what happens after one application of the Ip operator, but as can be seen from the resulting configuration, further changes would occur. In the configuration on the right, a and c are both strongly influenced to believe p, and so a further application of Ip would cause them to revise their beliefs, returning them to their previous doxastic states. What happens if we continue in this way? Well, in this case, there is no further change to agents b and c, so after only one more application of Ip, all four agents unanimously believe that p. Social influence will cause no further changes.

We'll say that a community is *stable* if the operator Ip has no effect on the doxastic states of any agent in the community. Unanimity within the community is sufficient for stability but not necessary, as is shown below:

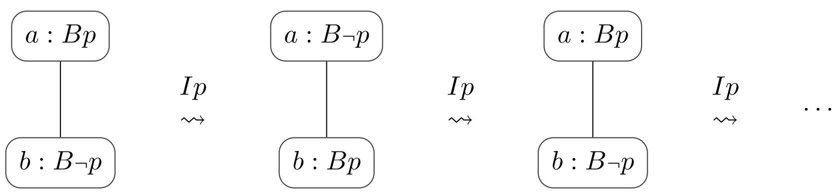
Example 2:



In this network, which consists of just one community, no agent is subject to either strong or weak influence. For example, agent b, who is undecided about p, has three friends who believe p but one that disbelieves. On the very weak assumptions we are making about threshold influence, this is not enough to get her to change her mind. Also note that agent c's friends are unanimous in being undecided about p, but this has no effect on c's belief in p.

The community of Example 1 is not stable but becomes stable after one application of Ip. We will say that such communities are *becoming stable*. Not all communities are becoming stable. Those that never become stable will be said to be *in flux*. Here is an example:

Example 3:



In Example 3, each agent is strongly influenced by the other at each stage, and so revises her belief regarding p, alternately believing and disbelieving p.

Thus, a brief examination of the dynamics of influence show that all three possibilities can be realised: communities that are stable, those that are in flux, and those that are not stable but are becoming stable. Moreover, since distinct communities in a social network have no influence over each other, it is possible to have a network with several communities of a different dynamic type.

This observation raises some questions for the logic of of friendship and belief. Under what conditions is an agent part of a community that is of each of these types? Characterising stability is fairly easy, because it can be done locally. An agent who is under neither strong nor weak influence to believe p, and is also under neither strong nor weak influence to believe $\neg p$, will not change her mind. But also, an agent who already believes p, and is under strong or weak influence to believe p, will also remain unchanged. Reflection on the automaton is enough to convince us that the following condition is necessary and sufficient for an agent not to change her mind (assuming that strong influence implies weak influence):

$$\neg(B\neg p \wedge \mathsf{W}p) \;\wedge\; \neg(Up \wedge \mathsf{S}p) \;\wedge\; \neg(Up \wedge \mathsf{S}\neg p) \;\wedge\; \neg(Bp \wedge \mathsf{W}\neg p)$$

Under the assumption of threshold influence, this is equivalent to

$$\begin{array}{l}\neg(B\neg p \wedge F\neg B\neg p \wedge \langle F\rangle Bp) \wedge \\ \neg(\neg Bp \wedge \neg B\neg p \wedge FBp \wedge \langle F\rangle Bp) \wedge \\ \neg((\neg Bp \wedge \neg B\neg p \wedge FB\neg p \wedge \langle F\rangle B\neg p) \wedge \\ \neg(Bp \wedge F\neg Bp \wedge \langle F\rangle B\neg p)\end{array}$$

which is clearly only in the language of friendship and belief. A community is stable when every agent in the community satisfies this condition. Characterising the state of being in flux (or, equivalently, becoming stable) is a little harder. [14] contains a theorem that shows how to do this for the preference dynamics studied there. Here, we will conjecture that a similar property holds for belief, namely, that a community (of at least two agents) is in flux if and only if every agent in the community satisfies the condition

$$(FBp \wedge FFB\neg p) \;\vee\; (FB\neg p \wedge FFBp)$$

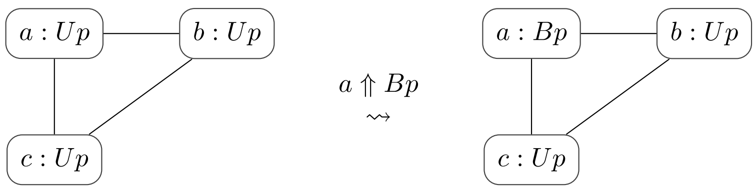
In particular, if there is any agent in the community in state Up, then the community is becoming stable, if not stable already.

This highlights the ease of attaining stability. In the next few sections we will study ways to break or introduce stability in a community, and how beliefs may propagate within a community or from one community to another as a result of influence.

2.1 Private belief change

Agents may change their minds for many reasons other than the influence of their friends' opinions. This raises the question of if and how such changes are propagated to other members of the community. A very coherent community may resist all such changes, ensuring that any agent who changes her mind unilaterally, will soon be brought back into conformity. On the other hand, a less coherent community, may be highly affected by the change, going into flux or even following the agent who changed her mind into a new stable configuration. We will examine some of the possibilities here starting with a single agent deciding to believe that p in a unanimously undecided community:

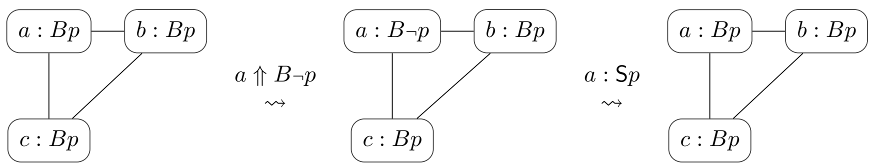
Example 4:



In the initial stable configuration, there are three agents, all of whom are friends and all undecided. Now, agent a changes her belief to Bp, which we denote in the diagram by '$a \Uparrow Bp$'. The change is very limited, however, because this new configuration is also stable. After the change, agent a has no pressure to drop her belief, as all her friends are merely undecided. Those friends, b and c, are under weak influence to believe p, but this is not strong enough for them to change from their undecided state under threshold influence. The private belief change is therefore completely isolated.

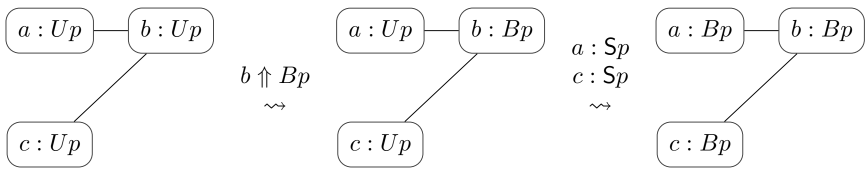
Unanimous belief within a community can be strong enough to resist private belief changes even further, as the next example shows.

Example 5:



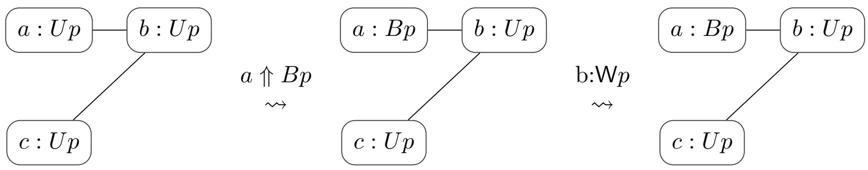
Here agent a first privately revises her belief in p to believe $\neg p$ but immediately reverses this change under strong influence from her friends (indicated by 'a : $\mathsf{S}p$' in the diagram). These two examples illustrate resistance to change in communities. Example 4 shows a passive resistance: the other agents are not influenced but tolerate the change. Example 5 shows more active resistance: the agent who changed her belief is forced back into conformity. More radical consequences of private belief change are possible, even with threshold influence, if we change the geometry of the social network.

Example 6:



For a community of undecided agents to be influenced by a private belief change, as in Example 6, the location of the agent who comes to believe p is critical. A peripheral agent will not succeed.

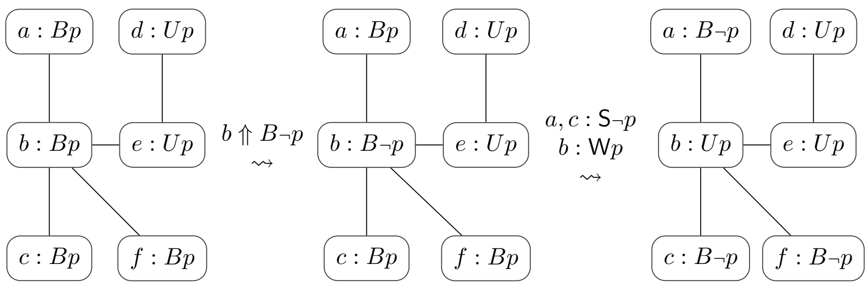
Example 7:



2.2 Aggregate belief change

We can get more of a sense of how private changes in belief affect the community by calcuating an aggregate opinion. This is a notoriously difficult thing to do well but here we will assume only that a sufficient condition for a community to have an aggregate belief in some proposition is that at least half of the members believe it and no member of the community disbelieves it, although some may be undecided. In such cases, we will say that the community has a *near unanimous opinion*. With even one agent privately changing her belief, even a near unanimous opinion can be overturned, as the following example shows:

Example 8:

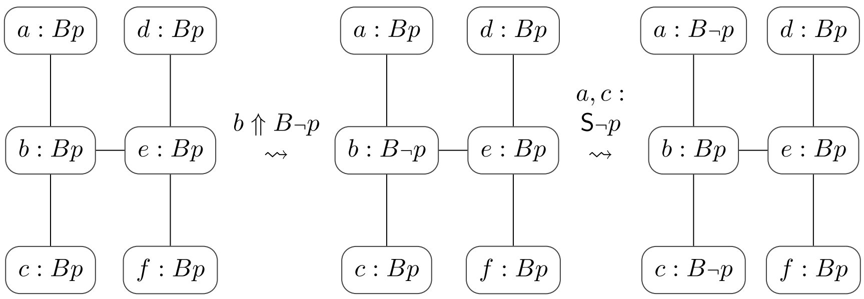


Here, the influence of centrally located agent b overturns the near unanimous opinion for p, achieving a group revision. Note that the presence of undecided agents d and e is crucial here. Without their stabilizing influence, the rest of the group would oscillate between believing and disbelieving.

In fact, whenever any two friends are of the same opinion, they will not be influenced to change their beliefs. Take, for instance, two friends a and b who both believe that p. The only way a will change his mind under strong influence is if all his friends believe $\neg p$, which will not happen as long as b continues to believe p. Even weak influence on a requires, at least, that his friends all do not believe p, for which b is again a counterexample. So long as b retains her conviction in p, a will be unaffected by social influence (of the '100% threshold' kind). But the situation is entirely symmetric, and so b will also be unaffected. The only way in which either will be influenced to change his or her mind is if one of them changes for some other reason (as a 'private' belief change.)

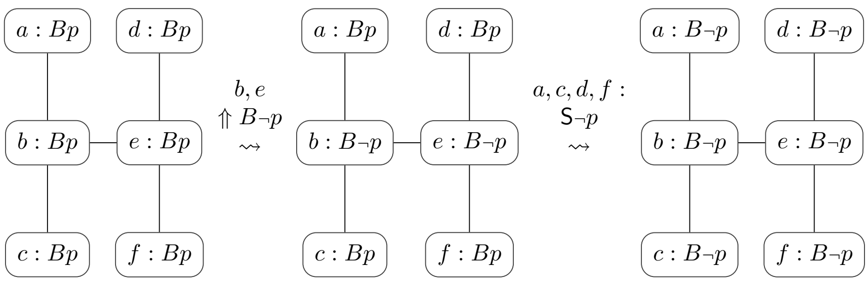
The stability of pairs makes it difficult for a single individual to affect large networks.

Example 9:



After one more iteration, the community will return to its initial unanimous belief. Yet, if two friends change their minds privately but simultaneously, we can get a total reversal of unanimous opinion:

Example 10:

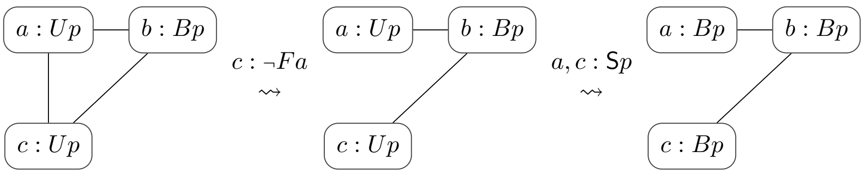


By allowing for private belief change, we can subject stable social opinions to a kind of 'stress test'. Very stable configurations will not be affected by an individual change of mind but less stable configurations will. Those configurations that are unaffected by two simultaneous private belief changes are even more stable. One could use this to define a measure of the 'resistence' of a community to changes of opinion. Given some method of aggregating the beliefs of members of a community (with respect to p), we can say that it is *n-resistant* if the aggregate opinion does not change as a result of any n members simultaneously changing their minds (in any way) with respect to p. Then the community of Example 8 is 0-resistant but not 1-resistant (with any reasonable aggregation mechanism) and that of Example 9 and 10 is 1-resistant but not 2-resistant.

2.3 Gaining and losing friends

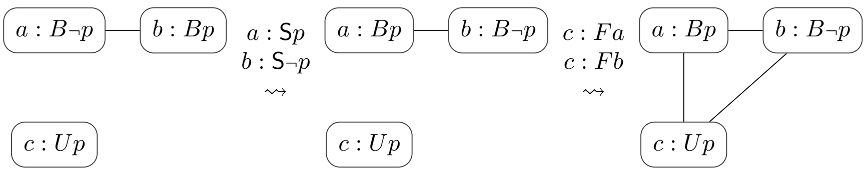
Changes to the social network can also lead to changes in aggregate opinion and so can be used to distinguish between more stable and less stable communities. The simplest of these occur with a single gain or loss of a friend.

Example 11:



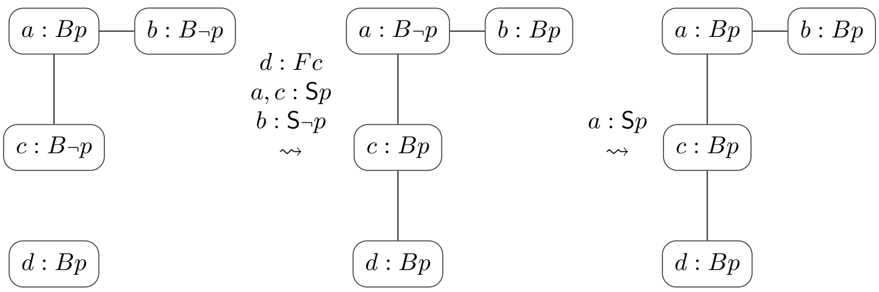
We start with a stable distribution of opinions among three mutual friends, with only one believer. One friendship is broken, putting the mutual friend into a position of greater influence. As well as changing the aggregate opinion, changes in the network can also change the dynamic status of communities, from stable to flux and vice versa

Example 12:



The oscillating pair of friends at the top is calmed when an indifferent agent joins their circle. In one more step, they will all become undecided. When the newcomer to a community in doxastic flux is a believer, the influence may be sufficient to convert the whole community:

Example 13:



Timing for newcomers is important, however. If the first agent that the newcomer met was of opposite opinion, he may be absorbed into the flux, unless, of course, he is part of another community that provides some stability.[9]

3 Alternatives to threshold influence

Our model of threshold influence is thoroughly egalitarian: when it comes to doxastic influence, all friends have equal power over us. One consequence of this assumption, as we saw above, is that pairs of friends of the same state of mind (believers, disbelievers or undecided) will never be influenced to change their beliefs. We will now consider a slightly different approach, whereby we think of the conditions for strong and weak influence as arising from ways of aggregating the opinions of our friends, who are ordered according to their relative power over us.

So, let's assume than our friends are (partially) ordered by a relation we will call 'better friend'. It is both irreflexive (no one is a better friend of mine that herself) and transitive (if a is a better friend than b, who is a better friend than c, then a is a better friend than c) but not necessarily linear: I may have two friends a and b neither of whom is a better friend than the other. This talk of 'better friend' is only a *façon de parler*; what we really mean by it is a kind of social power. To say that a is a better friend (of mine) than b means only that a has greater power to influence me than b.

For strong influence to revise our beliefs, we will suppose that only our best friends are consulted, but they must believe unanimously. This amounts to aggregating our best friends' opinions with a very strong requirement for aggregation, namely, unanimity. It is still very conservative but more liberal than the threshold condition we have been using so far, since we do not require anything of our wider circle of friends.

[9]For example, if d is initially friends with another believer, e, who is not connected to a, b or c, then d will be immune to change.

For weak influence to contract a belief, we will suppose that all our friends are consulted but that we aggregate their opinions in a way that gives priority to better friends. Specifically, we say that, on aggregate, our friends believe if for every disbelieving friend, we have a better friend who believes. This is a 'defeasibility' model of aggregation: the opinions of disbelieving friends are defeated by their betters.[10]

3.1 Ranked Influence

To capture this new model of social influence axiomatically, in full generality, we would need to distinguish between friends in terms of their power over us, introducing a binary operator B, such that $B\varphi\psi$ means 'for all my friends who φ, I have a better friend who ψ'. This presents some technical difficulties because it is not a normal modal operator.[11] Rather than tackle these difficulties here, we will introduce a simplifying assumption using the concept of *rank*. Our best friends are of rank one. Those who are the best of the remainder (when we remove our best friends) are of rank two. Those who are the best of the remainder (when we also remove our friends of rank two) are of rank three. And so on. The simplifying assumption is that having a higher rank is also sufficient for being a better friend: that if a is a higher ranked friend (for me) than b, then a is a better friend of mine than b.[12]

We assume that there are only a finite number N of agents in the social network, and so the lowest possible rank of friends is $N-1$.[13] We will therefore introduced new symbols $F_1, \ldots, F_{N-1}$, with F_i meaning 'all of my friends of rank i'. Thus we have

$$F\varphi \quad \leftrightarrow \quad \bigwedge_{i<N} F_i\varphi$$

We can now express ranked strong and weak influence with the following ax-

[10]This method of aggregation has been well-studied, although mainly with regard to preference rather than belief. Our ordering of friends is what is known as a 'priority graph' in [2], and the method itself is known as 'lexicographic aggregation'. To see the connection with dictionaries, think of a pair of words (of equal length) and the order they are listed. Word X comes before word Y just in case for every letter in Y that comes before the corresponding letter in word X (in alphabetic order), there is an earlier letter in Y that comes after the corresponding letter in word X. If the words are not of equal length, this definition can still be made to work by padding the shorter word with extra 'space' characters, which are considered to come before all the letters of the alphabet.

[11]The normal binary modal operator defined over the 'better friend' relation, $N\varphi\psi$, means 'for all my friends who φ, every better friend ψ'. But there is no way of defining B in terms of N.

[12]To see that this additional assumption is non-trivial, suppose I have one best friend, a, three other friends, b, c and d, with b a better friend than d, If b and c are incomparable, then neither is a better friend of mine than the other. But then a has rank 1, b and c have rank 2, and d has rank 3. This implies that c is a better friend of mine than d, which is an inference we could not make without the ranking assumption.

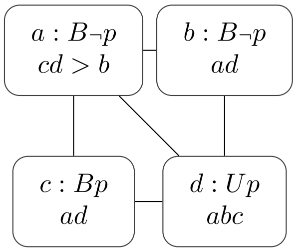
[13]Since friendship is assumed to be irreflexive, there must be at least two agents in order for there to be any friends at all.

ioms:[14]

$$
\begin{array}{lcl}
S\varphi & \leftrightarrow & F_1 B\varphi \wedge \langle F_1 \rangle B\varphi \\
W\varphi & \leftrightarrow & \bigwedge_{i<N} F_i(B\neg\varphi \rightarrow \bigvee_{j<i} \langle F_j \rangle B\varphi) \wedge \langle F \rangle B\varphi
\end{array}
$$

Consider the following social network:

Example 14:



Each node in the diagram now represents both what the agent believes and how she ranks her friends, represented by a simple list of the ranks. So, for example, a's best friends are c and d, with b in the second rank, and so a is weakly influenced to contract her disbelief in p. This is because her only friend who fails to believe p, namely b, is of second rank, and she has a best friend, c, who believes p. Thus under ranked influence, a will become undecided, whereas under threshold influence, a would not change her beliefs because her friend b shares her disbelief in p.

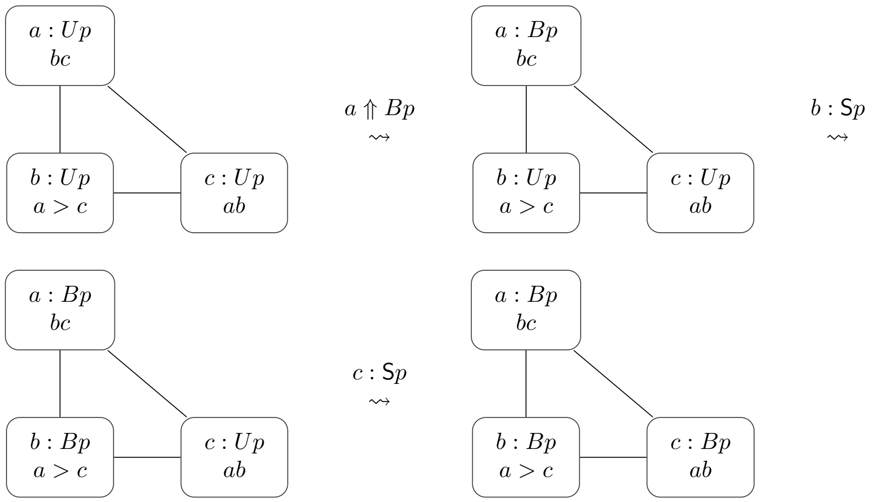
Ranked influence allows changes to spread within a community more easily, if conditions are right. In the next example, a single agent's new belief spreads to his community of previously undecided friends.

[14]The two axioms look superficially very different, but the first has an equivalent form that displays the difference more clearly:

$$
S\varphi \quad \leftrightarrow \quad \bigwedge_{i \leq N} F_i(\neg B\varphi \rightarrow \bigvee_{j<i} \langle F_j \rangle B\varphi) \wedge \langle F \rangle B\varphi
$$

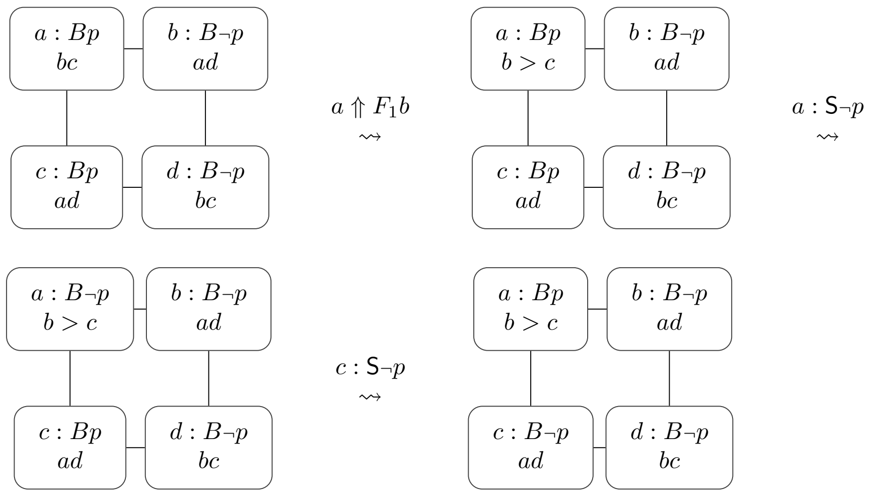
This expresses the apparently weaker condition that for every friend who does not believe φ, I have a better friend who does. But for this to be false, I must have a friend who doesn't believe φ and no better friend who does. But then either that friend is a best friend, or I have a best friend (and so a better friend) who does not believe φ, preventing strong influence.

Example 15:



Although the absolute stability of pairs of likeminded friends under threshold influence is disrupted when we move to ranked influence, paired best friends are still immune to change. But now we can also consider changes to the ranking of a given agent's friends, as a refinement of the actions of gaining and losing friends considered earlier.

Example 16:



Here, we represent the operation of promoting b to (sole) best friend by $\Uparrow F_1 b$. The new asymmetry allows b's disbelief to spread to the rest of the community. One could think of many other operations on rankings that could be studied.

For example, one could demote a friend, or promote/demote to a specific rank.

3.2 Believed Reliability

When the 'better friend' relation is re-interpreted in other more doxastically relevant ways, such as the relation of 'regarded as having more expertise than' or 'taken as a better authority than', the subjective component of an agent's rankings becomes evident. This raises the possibility of another way of modelling the relative power of other agents to affect us: to consider a binary relation of 'being more reliable than' between agents and then express the conditions for social influence in terms of which agents within one's community one *believes* to be more reliable. As in ranked influence, strong influence would require unanimity between those one takes to be most reliable, and weak influence to believe p would require any agent who disbelieves p to be defeated by an agent who believes p and whom one believes to be more reliable. Writing L for the operator 'every more reliable friend', we could then axiomatise this new notion of social influence, which we dub 'reliable influence', as follows:[15]

$$\begin{array}{lcl} S\varphi & \leftrightarrow & BF(\neg B\varphi \to \langle L\rangle B\varphi) \\ W\varphi & \leftrightarrow & B(F(B\neg\varphi \to \langle L\rangle B\varphi) \wedge \langle F\rangle B\varphi) \end{array}$$

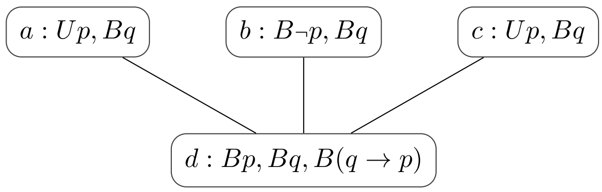
We will not go into the details of this approach here, merely noting that one would probably require some additional principles about the transparency of friendship, ensuring that one's beliefs about who is and who is not a friend are guaranteed to be true. It is enough for our present purposes to have shown that the possibilities for exploration of different conceptions of social influence are numerous, and that there are plenty of opportunities for representing these difference logically.

4 Plausibility influence

The focus of the preceding sections has been the dynamics of belief change in a community. In the interest of providing clear examples we have considered social influence with respect to a single proposition. But in doing so, we have ignored a very important aspect of a person's beliefs, namely their interdependence. Changing one belief may well affect other beliefs. So a natural question to ask is whether the order matters when it come to calculating social influence. The simple answer is yes.

[15] The form of the axiom for strong influence adapts the alternative axiom of the ranked version given in Footnote 14.

Example 17:



In the depicted situation, how is d to change her beliefs? We will see that the outcome depends on whether we consider influence with respect to p and then q or vice versa. First, considering the proposition p, because all of her friends either believe $\neg p$ or are undecided on p, and assuming threshold influence, we conclude that she will contract and so come to be undecided about p. But since she also believes q and $q \to p$, merely removing p from her belief set would not be successful; she would also have to remove either her beliefs that q or her belief that $q \to p$. Now, suppose that her belief that $q \to p$ is more entrenched than her belief that q, so she removes q from her belief set.[16] Yet all of d's friends also believe q and so, under strong influence, she will revise her beliefs so that she believes q, and then, believing $q \to p$, she will again believe p. The result of considering influence in the order p, q is therefore that d's beliefs are unchanged. But in the opposite order, she will cease to believe both p and q.

This is an uncomfortably strange result, and to address it, we will move away from those approaches to belief revision that take it to be an operation defined on the propositional contents of beliefs. This is the dominant tradition in the literature but there are alternatives. The one we consider here is the tradition of [4] and [13] in which an agent's beliefs are taken to be supervenient on her judgements regarding the plausibility of specific outcomes. As we will see, this provides us with a solution to the problem of multiple issues, and enables us to make some interesting distinctions in the social setting, but comes with its own challenges.

Given a fixed domain W of possible outcomes, we will consider each agent's judgements regarding the relative plausibility of elements of W. For u and v in W, we write $u \leq_a v$ to mean that a judges v to be at least as plausible as u. Importantly, it is possible for this relation to fail to be antisymmetric: two outcomes may be regarded as equally plausible. Also, the relation may fail to be total: there may be outcomes u and v about which the agent has no judgement regarding their relative plausibility: $u \not\leq_a v$ and $v \not\leq_a u$. Thus, one fundamental change from our previous model is that there are now four (rather than three) relevant possible states of an agent: agent a may find v strictly more plausible than u ($u \leq_a v$ and $v \not\leq_a u$) or vice versa, or may regard them as equally plausible ($u \leq_a v$ and $v \leq_a u$) or have no view at all ($u \not\leq_a v$ and $v \not\leq_a u$).

There are various ways in which one might think of plausibility judgements as determining beliefs but the dominant idea (from [4] and [13]) is that an agent

[16]For more on entrenched belief change, see [10] and [11].

a believes that p just in case p holds in all the outcomes that are maximally plausible for a. An outcome is 'maximally plausible' just in case there is no other outcome that the agent judges to be strictly more plausible. Although some suggestions have been made about how to model belief revision in this framework, we will consider the separate but related question of how to model revision of plausibility judgements themselves. In the first instance, this is much more straightforward than revising her beliefs. If she wishes to revise so that she judges w_2 to be at least as plausible as w_1, she should revise to

$$\leq_a \cup \{\langle w_1, w_2 \rangle\}$$

In other words, she should regard v as at least as plausible as u iff she previously regarded v as at least as plausible as u or $u = w_1$ and $v = w_2$.[17] Likewise, if she wishes to revise so that she judges w_2 to be not at least as plausible as w_1, she should revise to

$$\leq_a \setminus \{\langle w_1, w_2 \rangle\}$$

Just as in our earlier model, we must also specify the conditions under which these revisions are made. This time, we will consider only the simplest possible proposal: that an agent revises her plausibility judgements (in a positive or negative direction) iff all her friends are unianimous. So, if they all take v to be at least as plausible as u, so does she, and if they all take v not to be at least as plausible as u, nor does she. We will call this *plausibility influence* and write the corresponding operator as $\mathcal{I}$.[18] Now, let's focus on a particular couple of outcomes, u (the 'left' one) and v (the 'right' one). The action of plausibility influence is characterised by the following automaton:

$$\begin{tikzcd}
\boxed{u \leq v} \arrow[rr, bend left, "\dot{F}(u \not\leq v)"] & & \boxed{u \not\leq v} \arrow[ll, bend left, "\dot{F}(u \leq v)"]
\end{tikzcd}$$

(where $\dot{F}\varphi$ is an abbreviation for $(F\varphi \wedge \langle F \rangle \varphi)$, i.e. the version of the universal quantifier that takes it to have existential import.) Of course, this is only half

[17] This proposal raises certain problems, especially concerning the transitivity of plausibility judgements. We will address these below.

[18] More precisely, plausibility influence is the operation that transforms the plausibility judgements of all agents in such a way that agent a deems v to be at least as plausible as u iff the pair $\langle u, v \rangle$ is in the set

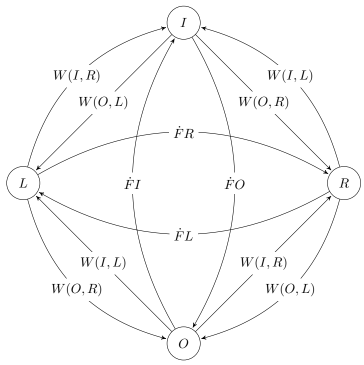
$$\leq_a \cup \bigcap_{a \asymp b} \leq_b \setminus \bigcap_{a \asymp b} \not\leq_b$$

where $x \asymp y$ means that x is friends with y. Note that the order in which the operations of adding and subtracting from the relation are performed is not important because, with at least one friend, it can never be that all my friends both do and do not regard v as at least as plausible as u.

the story, the other half of which is given by swapping u and v. As mentioned above, an agent can have one of four attitudes with respect to u and v, which we will label as R (for 'right') v is strictly more plausible than u, L (for 'left') u is strictly more plausible than v, I (for 'impartial') u and v are equally plausible, and O (for 'no opinion'). In these terms $u \leq v$ is $(I \vee R)$, $u \not\leq v$ is $(O \vee L)$, $v \leq u$ is $(I \vee L)$, and $v \not\leq u$ is $(O \vee R)$. The dynamics of the two parts of the comparison can therefore be represented as the two automata

$$\begin{tikzcd} I \vee R \arrow[r, bend left, "\dot{F}(O \vee L)"] & O \vee L \arrow[l, bend left, "\dot{F}(I \vee R)"] \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} I \vee L \arrow[r, bend left, "\dot{F}(O \vee R)"] & O \vee R \arrow[l, bend left, "\dot{F}(I \vee L)"] \end{tikzcd}$$

whose product completely describes the dynamics of plausibility influence with respect to the four states L, R, I and O:



where $W(\alpha, \beta)$ is the condition $(\dot{F}(\alpha \vee \beta) \wedge \neg \dot{F}\beta)$.[19] This analysis of the dynamics shows that little has changed from our earlier models, in the sense that friends of the same type (L, R, I, or O) will be immune to influence from others; there are the typical unstable alternations of the form X — Y — X where X and Y are agent 'types'; and a characterisation of stability could be obtained using the methods of [14].

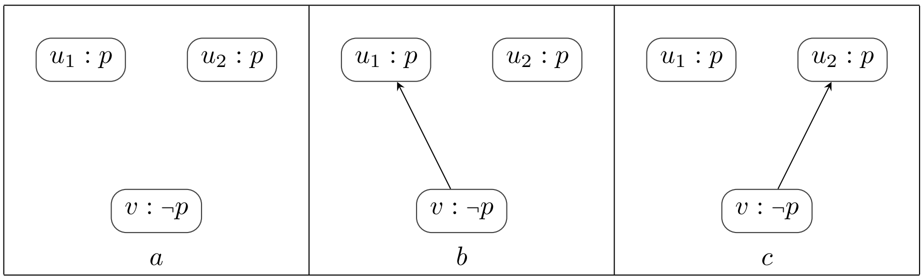
4.1 A Finer-grained Dynamics

The novelty of the plausibility approach to social influence is its fine-grained analysis of belief dynamics. Although it is still deterministic at the level of

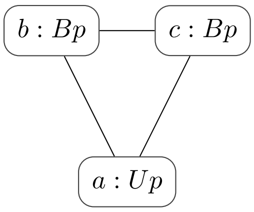
[19] As before, we will save cluttering our diagrams by assuming that if none of the conditions for a transition apply then the automaton stays in its current state.

plausibility, unlike the revision/contraction approach, it is *not* deterministic at the level of beliefs. For example, suppose agents b and c are friends who both believe p. In our previous model, this makes them invulnerable to influence with respect to this belief, and if there is a third agent a who is friends with both of them (and no one else), she will be strongly influenced to believe p also, no matter what her initial view. But with plausibility influence this is no longer the case.

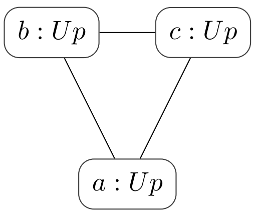
Suppose there are three possible outcomes, u_1, u_2 and v with p true at each of u_1 and u_2 but not at v. And suppose that agent b regards u_1 as strictly more plausible than v, agent c regards u_2 as strictly more plausible than v but that these are the only judgements they make. In particular, agent a makes no judgements at all. This is depicted below:



Agents b and c agree that u_1 and u_2 are the only maximally plausible outcomes, and so believe that p. Their friend a also allows that outcome v is maximally plausible and so is undecided about p. Thus we have the following configuration:



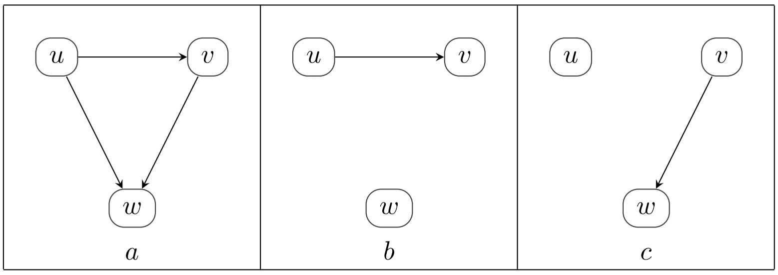
But after plausibility influence, agent b drops her judgement that u_1 is more plausible than v because it is not supported by either friend. Likewise, agent c drops her judgement that u_2 is more plausible than v, making all three agents converge to a's initial view, and so the new configuration is



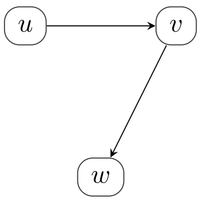
We can interpret this as capturing, to some extent, the influence of reasons rather than mere beliefs. On this interpretation, the agent's reasons for believing are given by her plausibility judgements; it is those that are influenced by her social environment. Influence on her beliefs is a secondary matter.

4.2 Transitivity

One problem with plausibility influence is that it does not work very nicely with the rather natural requirement that plausibility judgements are transitive. For example, suppose instead that the three friends make the following plausibility judgements about outcomes u, v and w:



Each of these satisfies transitivity. But, under plausibility influence, each friend will change to the paradigmatically intransitive set of plausibility judgements given by



Put briefly, the operation of plausibility influence fails to preserve transitivity. The are several ways to respond to this. One is simply to take the transitive closure of each agent's plausibility relation after calculating the effect of social influence, perhaps interpreted as an act of self-critical evaluation. This would make it marginally more difficult for an agent to be influenced by his peers. A slight drawback is that it becomes more difficult to interpret the dynamics as operating on plausibility judgements as reasons for beliefs: taking w to be more plausible than u is not a reason *in addition* to taking w to be more plausible than v and v to be more plausible than u. A finer-grained dynamics of plausibility judgements as reasons for belief, of the kind suggested above, would have to make a distinction between primary judgements and those that are inferred by transitivity. A second response is to modify the definition of belief slightly, so that it is the transitive closure of the plausibility relation, for each agent, that determines the agent's beliefs.

5 Conclusion

We have conducted a fairly open-ended survey of some of the possibilities for modelling the way in which people's beliefs are influenced by their social relations. Unlike models offered by social psychologists and sociologists, our aim is not descriptive but normative. We are interested in the relationship between norms of belief revision that may be adopted by members of a community and the resulting dynamic properties of the distribution of beliefs across that community. Nonetheless, we have seen, at a qualitative level, that many aspects of social belief change that we see in real communities can be obtained from even a very simplistic model, which we called 'threshold influence'. In particular, we focussed on the question of what makes the beliefs of a community stable under the dynamics of influence itself and various 'stress tests' such as unilateral changes of belief by individuals within the community and changes to the social network itself. We also considered refinements and alternatives to the 'threshold' model. The most significant alternative was to move to consideration of plausibility judgements rather than mere beliefs. We showed first that some such change is mandated by difficult problems with belief-based dynamics related to the need to decide on an order in which different beliefs are considered. Secondly, we showed that the resulting plausibility-based account results in a dynamical system that is non-deterministic at the level of beliefs. Nonetheless, the plausibility-based account we considered lacks certain intuitively desirable features, such as the preservation of the transitivity of plausibility judgements. With respect to each of the above points, there is a lot of opportunity for future work.

References

[1] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] H. Andréka, M. Ryan, and P-Y. Schobbens. Operators and laws for combining preferential relations. *Journal of Logic and Computation*, 12:12–53, 2002.

[3] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, Cambridge, Mass., 2001.

[4] J. Burgess. *Basic Tense Logic*, volume 2 of *Handbook of Philosophical Logic*, chapter 2, pages 89–133. Dordrecht: D. Reidel, 1984.

[5] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, pages 57–66. ACM Press, 2002.

[6] John French. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.

[7] Nigel Gilbert and Klaus Troitzsch. *Simulation for the Social Scientist.* Open University Press, 2 edition, February 2005.

[8] Matthew O. Jackson. *Social and Economic Networks.* Princeton University Press, August 2008.

[9] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings the 9th Acm International Conference*, pages 137–146. ACM Press, 2003.

[10] A. Nayak, P. Nelson, and H. Polansky. Belief change as change in epistemic entrenchment. *Synthese*, 109(2):143–174, 1996.

[11] Hans Rott. Basic entrenchment. *Studia Logica*, 73:257–280, 2003.

[12] J. Seligman, F. Liu, and P. Girard. Logic in the community. In Mohua Banerjee and Anil Seth, editors, *Logic and Its Applications - 4th Indian Conference, ICLA 2011, Delhi, India, January 5-11, 2011. Proceedings*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188, 2011.

[13] Frank Veltman. *Logics for Conditionals.* PhD thesis, University of Amsterdam, 1985.

[14] Liang Zhen and Jeremy Seligman. A logical model of the dynamics of peer pressure. *Electronic Notes in Theoretical Computer Science*, 278:275–288, 2011.