

A Logical Model of the Dynamics of Peer Pressure

Liang Zhen¹

*Institute of Logic and Intelligence
Southwest University
Chongqing, China*

Jeremy Seligman²

*Department of Philosophy
University of Auckland
New Zealand*

Abstract

Following the general programme of [6], we investigate the effect of social relationships on the dynamics of preference change within a community. Specifically, we are interested in the phenomenon of ‘peer pressure’, according to which a person’s preferences are changed in response to the preferences of a ‘peer group’. This involves both aggregation of preferences, to determine the group’s preferences and preference change. We propose a simple model of peer pressure which is still sufficiently non-trivial to display some interesting dynamics, and show how the stable configurations can be expressed logically.

Keywords: preference logic, logic in the community, aggregation, social dynamics, hybrid logic

1 An Example

A community of teenage girls is divided between two current fashions. Some strictly prefer *A*, others strictly prefer *B*, others are *I* indifferent and yet others are conflicted *O*. They are influenced by their friends’ choices in the following way. If all of their friends strictly prefer one of the styles then they will too. We’ll call this *strong suggestion*. Even if some of their friends are indifferent, if the rest strictly prefer one style, they will still be influenced, but not as strongly. We’ll call this *weak suggestion*. In this case, they will change their preference so that they regard

¹ Email: soul21@163.com

² Email: j.m.seligman@gmail.com

the style preferred by their non-indifferent friends so that it is at least as good as the other one. This means that if they are conflicted they will resolve the conflict in favour of their friend’s choice, and if they prefer the other style, they will become indifferent. In all other cases the girls don’t change their opinion.

The resulting dynamics has some interesting features. Representing the friendship relation as a graph we see the following looping patterns:

$$A - B \rightsquigarrow B - A \rightsquigarrow A - B \rightsquigarrow B - A \rightsquigarrow \dots$$
$$B \begin{smallmatrix} \nearrow A \\ \searrow A \end{smallmatrix} \rightsquigarrow A \begin{smallmatrix} \nearrow B \\ \searrow B \end{smallmatrix} \rightsquigarrow B \begin{smallmatrix} \nearrow A \\ \searrow A \end{smallmatrix} \rightsquigarrow A \begin{smallmatrix} \nearrow B \\ \searrow B \end{smallmatrix} \rightsquigarrow \dots$$

If one of the friends is indifferent, however, the pattern stabilizes:

$$\begin{array}{ccccccc} A - B - A - B & \rightsquigarrow & B - A - B - I & \rightsquigarrow & A - B - I - B & \rightsquigarrow & B - I - B - B \\ | & & | & & | & & | \\ A - B - I & & B - I - B & & B - B - B & & B - B - B \end{array}$$

A configuration $I \begin{smallmatrix} \nearrow A \\ \searrow B \end{smallmatrix}$ with an indifferent mutual friend is stable and even when the indifferent friend is out on a wing of the group, the pattern stabilizes quickly to that of the mutual friend:

$$B \begin{smallmatrix} \nearrow A \\ \searrow I \end{smallmatrix} \rightsquigarrow I \begin{smallmatrix} \nearrow B \\ \searrow B \end{smallmatrix} \rightsquigarrow B \begin{smallmatrix} \nearrow B \\ \searrow B \end{smallmatrix}$$

When all three are friends, however, indifference spreads: $I \begin{smallmatrix} \nearrow A \\ \searrow B \end{smallmatrix} \rightsquigarrow I \begin{smallmatrix} \nearrow I \\ \searrow I \end{smallmatrix}$ Indifference is not so easily shaken, as $I \begin{smallmatrix} \nearrow I \\ \searrow A \end{smallmatrix}$ is stable. But those who are conflicted are slightly more easy to influence:

$$O \begin{smallmatrix} \nearrow I \\ \searrow A \end{smallmatrix} \rightsquigarrow A \begin{smallmatrix} \nearrow I \\ \searrow A \end{smallmatrix} \rightsquigarrow A \begin{smallmatrix} \nearrow A \\ \searrow A \end{smallmatrix}$$

Two friends of the same type will stand in solidarity, no matter what their environment:

$$\begin{array}{ccccc} & B & - & I & - & I & - & A \\ & | & & & & & & | \\ & B & - & O & - & O & - & A \end{array}$$

2 Preference Logic in the Community

The general framework we adopt to discuss this phenomenon is from [6]. Specifically, we will use a version of ‘facebook logic’, in which a community is modelled by a symmetric relation of ‘friendship’ which may vary across a space of possibilities. Although real communities involve much more complex social relations, ‘friendship’ alone is sufficient to demonstrate interesting dynamic behaviour.

Our models, then, will have a set A of *agents*, a set W of *possible states* and the following relations:

- For each $w \in W$, a symmetric and irreflexive relation \sim_w on A , which we interpret as friendship: $a \sim_w b$ means that in state w , a is friends with b . The *community* of a is the set of agents to whom a is connected by a chain of friends, and this is represented by \sim^* , the transitive closure of \sim .
- For each $a \in A$, a relation \leq_a on W , which we interpret as preference: $u \leq_a v$ means that for a , state v is at least as good as state u .

A *community preference frame*, then, is a structure $F = \langle W, A, \sim, \leq \rangle$. We allow for the case that our agent’s preferences are not fully rational, in the sense that \leq_a may not be transitive. In the case that it is, we will say that F is *transitive*.

We reason about such frames using a modal language

$$\varphi ::= i \mid n \mid p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid F\varphi \mid F^*\varphi \mid P\varphi \mid U\varphi$$

where $i \in \mathbf{Snom}$, a set of *state nominals*, $n \in \mathbf{Anom}$, a set of *agent nominals*, $p \in \mathbf{Prop}$, a set of *propositional variables*. The nominals are included in the language so that we can reason about particular agent’s preferences concerning particular states. A formula φ in this language is an indexical proposition, which is true or false in a state w about an agent a . In particular, the modal operators are interpreted as follows:

- $F\varphi$ means ‘all my friends, φ ’
- $F^*\varphi$ means ‘everyone in my community, φ ’
- $P\varphi$ means ‘all states in which φ holds are least as good as the current one (according to me)’
- $U\varphi$ means ‘in all states, φ ’

More precisely, given a *model* $M = \langle F, V \rangle$ where $V: \mathbf{Prop} \cup \mathbf{Snom} \cup \mathbf{Anom} \rightarrow \mathcal{P}W \times A$ is a *valuation* function such that $V(i)$ is $\{\underline{i}\} \times A$ for each $i \in \mathbf{Snom}$ and $V(n)$ is $W \times \{\underline{n}\}$ for each $n \in \mathbf{Anom}$, we define

$M, w, a \models p$	iff	$\langle w, a \rangle \in V(p)$
$M, w, a \models i$	iff	$\langle w, a \rangle \in V(i)$ (iff $w = \underline{i}$)
$M, w, a \models \neg\varphi$	iff	$M, w, a \not\models \varphi$
$M, w, a \models (\varphi \wedge \psi)$	iff	$M, w, a \models \varphi$ and $M, w, a \models \psi$
$M, w, a \models F\varphi$	iff	$M, w, b \models \varphi$ for each $b \sim_w a$
$M, w, a \models F^*\varphi$	iff	$M, w, b \models \varphi$ for each $b \sim_w^* a$
$M, w, a \models P\varphi$	iff	$w \leq_a v$ for each $v \in W$ such that $M, v, a \models \varphi$
$M, w, a \models U\varphi$	iff	$M, v, a \models \varphi$ for each $v \in W$

Notice in particular that P is *not* a normal modal operator. It is a ‘window’ operator in the sense of [1] (pp. 426–427). We need this to define preference between arbitrary propositions:

$(\varphi \leq \psi)$	$= U((\varphi \wedge \neg\psi) \rightarrow P(\psi \wedge \neg\varphi))$	weak preference for ψ over φ
$(\varphi < \psi)$	$= (\varphi \leq \psi) \wedge \neg(\psi \leq \varphi)$	strict preference for ψ over φ
$(\varphi = \psi)$	$= (\varphi \leq \psi) \wedge (\psi \leq \varphi)$	indifference concerning ψ and φ
$(\varphi \neq \psi)$	$= \neg((\varphi \leq \psi) \vee (\psi \leq \varphi))$	conflicted preferences about ψ and φ

In other words, I regard ψ as at least as good as φ iff every ψ -but-not- φ -state is at least as good as every φ -but-not- ψ -state. My preferences concerning states that are both φ and ψ or neither φ nor ψ are irrelevant. And I *prefer* ψ to φ iff ψ is at least as good as φ but φ is not at least as good as ψ . The latter hold just in case there is at least one ψ -but-not- φ -state that is strictly better than all of the φ -but-not- ψ -states, but in conjunction with the first clause, this is equivalent to every ψ -but-not- φ -state being strictly better than every φ -but-not- ψ -states.

Two special cases are worth mentioning from the outset. Firstly, the definition is consistent with preferences between particular states: $(i \leq j)$ holds for agent a iff $\underline{i} \leq_a \underline{j}$. Secondly, we can represent a *weak preference for* φ as $(\neg\varphi \leq \varphi)$. Since φ and $\neg\varphi$ are nowhere satisfied together, this is equivalent to $U(\neg\varphi \rightarrow P\varphi)$. This holds when all the φ -states are at least as good as all the $\neg\varphi$ states. A *preference for* φ , represented by $(\neg\varphi < \varphi)$ holds when at least one φ -states is strictly preferred to all the $\neg\varphi$ -states.

Many definition of preference have been offered in the literature, but we believe that this one captures the simplest concept of rational preference, namely preference *ceteris paribus*. That is, if we assume that all of the states are equivalent ‘other things being equal’ then one should prefer ψ to φ under the specified condition.³

Some further abbreviation are useful:

³ Note that \leq is reflexive, in the sense that $(\varphi \leq \varphi)$ is valid.

$$\begin{aligned}
\langle F \rangle \varphi &=_{\text{df}} \neg F \neg \varphi & M, w, a \models \langle F \rangle \varphi &\text{ iff } M, w, b \models \varphi \text{ for some } b \sim_w a \\
\langle P \rangle \varphi &=_{\text{df}} \neg P \neg \varphi & M, w, a \models \langle P \rangle \varphi &\text{ iff } M, v, a \not\models \varphi \text{ for some } v \not\leq_a w \\
\langle U \rangle \varphi &=_{\text{df}} \neg U \neg \varphi & M, w, a \models \langle U \rangle \varphi &\text{ iff } M, c, a \models \varphi \text{ for some } v \in W \\
@_i \varphi &=_{\text{df}} U(i \rightarrow \varphi) & M, w, a \models @_i \varphi &\text{ iff } M, \underline{i}, a \models \varphi \\
i = j &=_{\text{df}} U(i \rightarrow j) & M, w, a \models i = j &\text{ iff } \underline{i} = \underline{j}
\end{aligned}$$

3 Preference and preference change

To model these changes we consider two operators $[X \leq Y]$ and $[X < Y]$ which upgrade the agent's preferences in response to the perception that other agents have the corresponding preferences regarding X and Y , which are subsets of W . We will be most interested in the case where X and Y are denotations of formulas, i.e., sets of the form $[\varphi] = \{w \in W \mid M, w, a \models \varphi\}$. But the definition of these operators is easier to grasp in a more abstract setting.

For any relation \leq on W , the operators are defined as follows. First let $(X \leq Y)$ be the set of $\langle u, v \rangle$ such that $u \in X$ and $v \in Y$. And note that $M, w, a \models (\varphi \leq \psi)$ iff $([\varphi] \leq [\psi]) \subseteq \leq_a$. Now we will first assume that X and Y are disjoint. Then

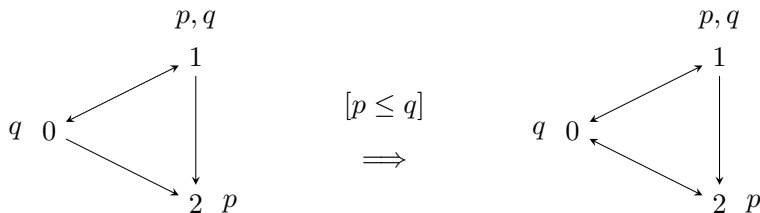
$$\begin{aligned}
[X \leq Y](\leq) &\text{ is } \leq \cup (X \leq Y) \\
[X < Y](\leq) &\text{ is } (\leq \cup (X \leq Y)) \setminus (Y \leq X)
\end{aligned}$$

In other words, $[X \leq Y]$ updates the agent's preference relation by *adding* links from all X -states to all Y -states, and $[X < Y]$ does the same but also *deletes* links from all Y -states to all X -states. In the case that X and Y are not disjoint we use $X \setminus Y$ and $Y \setminus X$ in place of X and Y :

$$\begin{aligned}
[X \leq Y] &= [(X \setminus Y) \leq (Y \setminus X)] \\
[X < Y] &= [(X \setminus Y) < (Y \setminus X)]
\end{aligned}$$

This limits the upgrade to those states on which X and Y disagree. A state that is in both X and Y or in neither of them should not be affected by the upgrade.

In both cases, these operations may not preserve the transitivity of \leq . For example,



When we only add links, in the case of $[X \leq Y]$ there is a way to repair any loss

of transitivity, by adding more links. There is a unique smallest transitive relation containing $[X \leq Y](\leq)$, namely its transitive closure. This can be defined in a local way, as the following lemma shows:

Lemma 3.1 *Let $(X \leq Y)^*$ be the set of $\langle u, v \rangle$ such that there is an $x \in X$ and a $y \in Y$ for which $u \leq x$ and $y \leq v$. Then $\leq \cup (X \leq Y)^*$ is the transitive closure of $[X \leq Y](\leq)$.*

Proof. Suppose $\langle u, v \rangle$ and $\langle v, w \rangle$ are in $\leq \cup (X \leq Y)^*$.

- (i) $u \leq v$. If $v \leq w$ then $u \leq w$. Otherwise, there is $x \in X$ and $y \in Y$ such that $v \leq x$ and $y \leq w$. But $u \leq v$ so $u \leq x$ and $y \leq w$ and so $\langle u, w \rangle \in (X \leq Y)^*$.
- (ii) $v \leq w$. Similarly.
- (iii) There are $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ such that

$$u \leq x_1 \text{ and } y_1 \leq v$$

$$v \leq x_2 \text{ and } y_2 \leq w$$

and so because of x_1 and y_2 , $\langle u, w \rangle \in (X \leq Y)^*$.

□

When we delete links, in the case of $[X < Y]$, the situation is more complicated. We would have to repair by deleting more links, and there need be no unique maximal way of doing this. This means that the operation of upgrading preferences so as to make them strict, while retaining transitivity is non-deterministic. For the remainder of the paper, we will consider only the case this non-transitivity-preserving but deterministic operator, so as to simplify the dynamics.⁴

In a model M , we write the result of upgrading agent a 's preferences with these operators as $[X \leq Y]_a M$ and $[X < Y]_a M$ respectively.⁵ We can then add a dynamic component to our logic, by adding unary operators

$$[\varphi \leq \psi] \text{ and } [\varphi < \psi]$$

with the following semantic conditions

$$M, w, a \models [\varphi \leq \psi] \theta \quad \text{iff} \quad [[\varphi] \leq [\psi]]_a M, w, a \models \theta$$

$$M, w, a \models [\varphi < \psi] \theta \quad \text{iff} \quad [[\varphi] < [\psi]]_a M, w, a \models \theta$$

As with many dynamic operators, the success of a preference upgrade is not guaranteed. For example, in the following model $[Pq < Pp] \neg (Pq < Pp)$:

⁴ Liu in [5], p.23-4, proposes a weaker upgrade operator that deletes links only in the case of indifference, and shows that it preserves transitivity. But this is not strong enough to ensure success, in our sense.

⁵ More precisely, $M = \langle W, A, \sim, \leq \rangle$ then $[X \leq Y]_a M = \langle W, A, \sim, \leq' \rangle$ where $\leq'_b = \leq_b$ for all $b \neq a$ and $\leq'_a = [X \leq Y](\leq_a)$. Similarly for $[X < Y]_a M$.

$$\begin{array}{ccc}
 p & 0 \leftrightarrow 1 & q \\
 & [Pq < Pp] & \\
 & \implies & \\
 p & 0 \leftrightarrow 1 & q
 \end{array}$$

But in the case that φ and ψ do not contain the operator P , there is no change to the valuation, the domain or the friendship structure, and so both $\varphi \leq \psi$ and $\varphi < \psi$ are valid.

4 Suggestion Dynamics

Consider the issue of whether α is better than β , on which members of a community may have different opinions. To make the following discussion a little easier to follow, we will fix these formulas now, although everything we say about them will be quite general, except that they must satisfy the four preference upgrade success conditions: $\alpha < \beta$, $\alpha \leq \beta$, $\beta < \alpha$ and $\beta \leq \alpha$. We will say that an agent is subject to *peer pressure* regarding the issue of α vs. β if she satisfies the following condition:

If (*Strong Suggestion*) all of her friends strictly prefer α to β (and she has at least one friend) then she will upgrade her preferences with $[\beta < \alpha]$, otherwise

if (*Weak Suggestion*) all of her friends regard β as at least as good as α and some of them strictly prefer α to β then she will upgrade her preferences with $[\beta \leq \alpha]$.

The conditions for the various types of suggestion can be expressed in our language as follows:

$S(\alpha, \beta)$	$(F(\alpha < \beta) \wedge \langle F \rangle(\alpha < \beta))$ my friends strongly suggest that β is better than α
$S(\beta, \alpha)$	$(F(\beta < \alpha) \wedge \langle F \rangle(\beta < \alpha))$ my friends strongly suggest that α is better than β
$W(\alpha, \beta)$	$(F(\alpha \leq \beta) \wedge \langle F \rangle(\alpha < \beta) \wedge \langle F \rangle(\beta \leq \alpha))$ my friends only weakly suggest that β is better than α
$W(\beta, \alpha)$	$(F(\beta \leq \alpha) \wedge \langle F \rangle(\beta < \alpha) \wedge \langle F \rangle(\alpha \leq \beta))$ my friends only weakly suggest that α is better than β
$N(\alpha, \beta)$	The negation of the disjunction of the above. my friends have no suggestion regarding α and β

The operation of suggestion transforms a model $M = \langle W, A, \sim, \leq, V \rangle$ to model $\#_{\alpha, \beta}M = \langle W, A, \sim, \#_{\alpha, \beta}(\leq), V \rangle$ where, for each $a \in A$,

$$\#_{\alpha, \beta}(\leq_a) = \begin{cases} [\alpha < \beta](\leq_a) & \text{if } M, a, u \models S(\alpha, \beta) \\ [\beta < \alpha](\leq_a) & \text{if } M, a, u \models S(\beta, \alpha) \\ [\alpha \leq \beta](\leq_a) & \text{if } M, a, u \models W(\alpha, \beta) \\ [\beta \leq \alpha](\leq_a) & \text{if } M, a, u \models W(\beta, \alpha) \\ \leq_a & \text{if } M, a, u \models N(\alpha, \beta) \end{cases}$$

We also add the suggestion operator to our language, with the obvious definition:

$$M, a, u \models \#_{\alpha, \beta}\varphi \quad \text{iff} \quad \#_{\alpha, \beta}M, a, u \models \varphi$$

So $\#_{\alpha, \beta}\varphi$ means that the agent is subject to peer pressure and after upgrading her preferences satisfies the description φ . It can be shown that $\#_{\alpha, \beta}\varphi$ is equivalent to the conjunction of the following formulas:

$$\begin{aligned} S(\alpha, \beta) &\rightarrow [\alpha < \beta]\varphi \\ S(\beta, \alpha) &\rightarrow [\beta < \alpha]\varphi \\ W(\alpha, \beta) &\rightarrow [\alpha \leq \beta]\varphi \\ W(\beta, \alpha) &\rightarrow [\beta \leq \alpha]\varphi \\ N(\alpha, \beta) &\rightarrow \varphi \end{aligned}$$

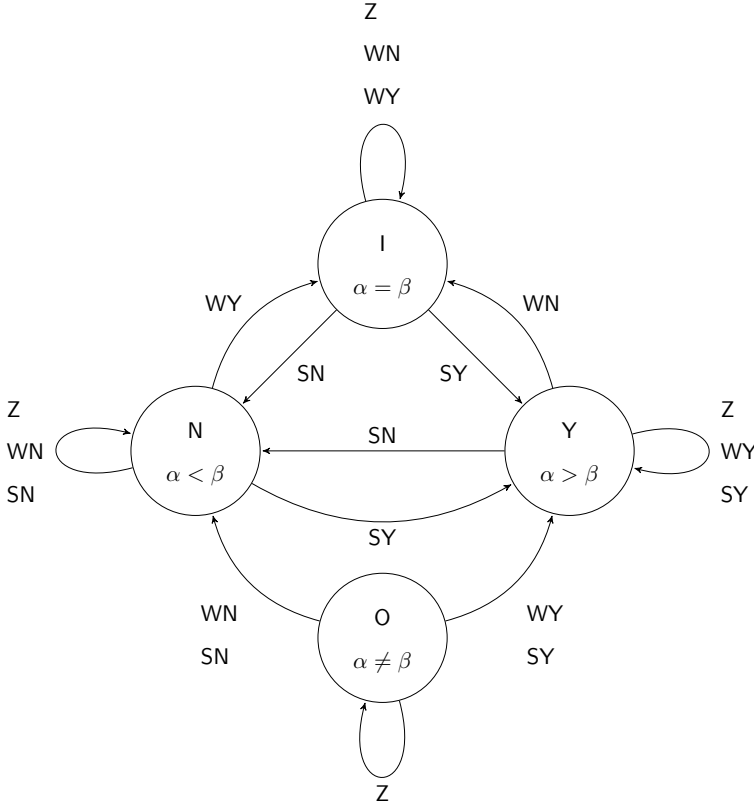
This is the basis for the dynamic behaviour observed earlier, and it is to apply our analysis of suggestion to such examples that we now turn. We will need some abbreviations for both the preference states of an agent concerning α and β and for the conditions under which they upgrade their preference in response to suggestion.

Y	$(\alpha > \beta)$	preference for α over β
N	$(\alpha < \beta)$	preference for β over α
I	$(\alpha = \beta)$	indifference concerning β and α
O	$(\alpha \neq \beta)$	conflicted preferences about β and α
SY	$FY \wedge \langle F \rangle Y$	my friends strongly suggest Y
SN	$FN \wedge \langle F \rangle N$	my friends strongly suggest N
WY	$F(I \vee Y) \wedge \langle F \rangle Y \wedge \langle F \rangle I$	my friends only weakly suggest Y

WN $F(I \vee N) \wedge \langle F \rangle N \wedge \langle F \rangle I$ my friends only weakly suggest N

Z $\neg(SY \vee SN \vee WY \vee WN)$ my friends have no suggestion

We can think of these as states of a finite state machine, depicted below.



A transition from state A to state B labelled with condition C represents the validity of the formula

$$(A \wedge C) \rightarrow \#_{\alpha, \beta} B$$

which is ensured by the definition of suggestion and the fact that α and β satisfy the preference upgrade success conditions. For example, if an agent is in state Y and her friends only weakly suggest N , so that WN is true, then after suggestion, she will be in state I , which is to say that she becomes indifferent.

5 Gaining Stability

The main result of this paper is to characterise the class of pointed models that stabilize under suggestion, in the manner illustrated in Section 1. An agent a in state u of model M has *stable preferences* iff for some preference state $\sigma \in \{Y, N, I, O\}$ for each $n \geq 0$,

$$\#_{\alpha, \beta}^n M, u, a \models \sigma$$

The model M is *stable* at u iff every agent has stable preferences at u . In other words, they are all completely fixed in their comparison of α and β and immune to suggestion from their friends. We have already seen a number of examples of stable models, and some that never become stable. In the middle are those models that change for a number of iterations before becoming stable. This is the class we will characterise. So, say that a model M *stabilizes* at u iff for some $n \geq 0$, $\#_{\alpha,\beta}^n M$ is stable at u .⁶ Likewise, we will say that an agent's preferences stabilize at u in M iff for some $n \geq 0$, a has stable preferences at u in $\#_{\alpha,\beta}^n M$.

Some formulas are very well behaved with respect to suggestion. They never change their truth value. We say that φ is an *invariant* of $\#_{\alpha,\beta}$ iff for any model M , state u and agent a ,

$$M, u, a \models (\varphi \rightarrow \#(\alpha, \beta)\varphi)$$

The central examples of invariant needed for our main theorem are as follows:

Lemma 5.1 *The following formulas are invariants of $\#_{\alpha,\beta}$:*

- | | |
|--|--|
| (1) $Y \wedge \langle F \rangle Y$ | (5) $N \wedge \langle F \rangle N$ |
| (2) $I \wedge \langle F \rangle I$ | |
| (3) $Y \wedge F(I \vee Y)$ | (6) $N \wedge F(I \vee N)$ |
| (4) $(Y \vee I) \wedge \langle F \rangle (Y \vee I)$ | (7) $(N \vee I) \wedge \langle F \rangle (I \vee N)$ |
| (8) $F^*((FY \wedge FFN) \vee (FN \wedge FFY))$ | |
| (9) $\langle F \rangle \top$ | |

Proof. (To make the presentation of this and subsequent proofs easier to read, we will abbreviate $\#_{\alpha,\beta}^n M, u, a \models \varphi$ to $n, a \models \varphi$. So to show invariance of φ , we have to show that $0, a \models \varphi$ implies $1, a \models \varphi$.)

- (1) Suppose $0, a \models Y \wedge \langle F \rangle Y$. Then $0, a \not\models SN$ and $0, a \not\models WN$. So from the state machine we see that a can only remain in state Y , i.e., $1, a \models Y$. But $0, a \models \langle F \rangle Y$ implies that a has a friend b in state Y and so $0, b \models Y \wedge \langle F \rangle Y$. This means that b must also remain in Y . Thus $1, b \models Y$, and so $1, a \models Y \wedge \langle F \rangle Y$.
- (2) The argument is similar to that for (1). From the state machine, we see that a and her friend must stay in state I .
- (3) Suppose $0, a \models Y \wedge F(I \vee Y)$. Since $0, a \models F(I \vee Y)$ we see that again $0, a \not\models SN$ and $0, a \not\models WN$ and so from the state machine, a must stay in Y . Now any friend b of a is either in state Y or in state I . In the first case, $0, b \models Y \wedge \langle F \rangle Y$ so by invariant (1), $1, b \models Y$. In the second case, since a is a friend of b , $0, b \models \langle F \rangle Y$ and so $0, b \not\models SN$. From the state machine, we see that b can only stay in I or move to Y . In either case, $1, b \models (Y \vee I)$ and so $1, a \models Y \wedge F(I \vee Y)$.

⁶ The constant reference to state u is required because in the community preference frame, friendship relations vary across states of the model.

- (4) Suppose $0, a \models (I \vee Y) \wedge \langle F \rangle (I \vee Y)$. Then there is a friend b of a such that both a and b are in states Y or N . If they are in the same, then they will keep these states by the invariant (1) or (2). And if they are different, without loss of generality, suppose $0, a \models Y$ and $0, b \models I$. Then $0, a \models \langle F \rangle I$ and so $0, a \not\models SN$ and must therefore either stay in state Y or move to state I , as shown in the state machine. Similarly, $0, b \models \langle F \rangle Y$ and so $0, b \not\models SN$ and must therefore either stay in state I or move to state Y . Either way, $1, a \models (I \vee Y) \wedge \langle F \rangle (I \vee Y)$.
- (8) Suppose $0, a \models F^*((FY \wedge FFN) \vee (FN \wedge FFY))$. Let b be a member of a 's community, i.e. $b \sim^* a$. Then also $0, b \models F^*((FY \wedge FFN) \vee (FN \wedge FFY))$ and, in particular, $0, b \models ((FY \wedge FFN) \vee (FN \wedge FFY))$. Then without loss of generality, suppose $0, b \models (FY \wedge FFN)$.

Let c be a friend of b . Then $0, c \models (Y \wedge FN)$. And because b is a friend of c , $0, b \models N$. But also $0, b \models F((FY \wedge FFN) \vee (FN \wedge FFY))$, so $0, c \models ((FY \wedge FFN) \vee (FN \wedge FFY))$. But $0, c \not\models (FY \wedge FFN)$ because $0, b \models N$ and so $0, c \models (FN \wedge FFY)$. Now $0, c \models SN$ because all his friends are N and so after suggestion, $1, c \models N$ and since c was chosen arbitrarily as a friend of b , we can conclude that $1, b \models FN$.

Now let d be a friend of c . Again $0, d \models ((FY \wedge FFN) \vee (FN \wedge FFY))$ because $0, b \models FF((FY \wedge FFN) \vee (FN \wedge FFY))$. But c is d 's friend and is in state Y so the second disjunct cannot hold and $0, d \models (FY \wedge FFN)$. So $0, d \models SY$ and $1, d \models Y$. Again d is an arbitrary friend of a friend of b , so $1, b \models FFY$.

Thus $1, b \models (FN \wedge FFY)$ and hence $1, b \models ((FY \wedge FFN) \vee (FN \wedge FFY))$. But b is an arbitrary member of a 's community, so $1, a \models F^*((FY \wedge FFN) \vee (FN \wedge FFY))$.

(9) is obvious and the remaining cases are Y - N symmetric versions of the others, and so require no further proof. \square

From these, we can show an important sufficient condition for stabilization:

Lemma 5.2 *If $M, u, a \models I$ then a 's preferences in M stabilizes at u for $\#_{\alpha, \beta}$.*

Proof. Suppose $0, a \models I$. Then from the state machine, we see that she can only stay in I or move either to N or to Y . So, either $n, a \models I$ for all n , and so her preferences stabilize, or after n iterations of the suggestion operator, she moves. Since the two possibilities are Y - N symmetric, we can assume she move to Y , so that $n+1, a \models Y$. But then, from the state machine, we see that $n, a \models SY$ and so $n, a \models (FY \wedge \langle F \rangle Y)$. Now consider any friend b of a . Then $n, b \models Y$. But a is friends with b , so $n, b \models \langle F \rangle I$, and so $n, b \not\models SN$. Then from the state machine, b can only stay in Y or move to I . So $n+1, b \models (I \vee Y)$. But then $n+1, a \models F(I \vee Y)$ and so $n+1, a \models (Y \wedge F(I \vee Y))$, which is an invariant by Lemma 5.1. \square

And now we are ready for the main theorem:

Theorem 5.3 M stabilizes at u for $\#_{\alpha,\beta}$ iff

$$M, u \models \neg(\langle F \rangle \top \wedge F^*((FY \wedge FFN) \vee (FN \wedge FFY)))$$

Proof. Let LOOP be the formula $F^*((FY \wedge FFN) \vee (FN \wedge FFY))$. If $M, u \not\models \neg(\langle F \rangle \top \wedge \text{LOOP})$ then we must show that M does not stabilize at u . So suppose that there is an agent a such that $M, u, a \models (\langle F \rangle \top \wedge \text{LOOP})$ and, for contradiction, that a 's preference are stable in $\#_{\alpha,\beta}^n M$ at u . But by Lemma 5.1, LOOP is an invariant and so $n, a \models \text{LOOP}$, which implies that

$$n, a \models ((FY \wedge FFN) \vee (FN \wedge FFY))$$

Without loss of generality, suppose $n, a \models (FY \wedge FFN)$. Also $\langle F \rangle \top$ is an invariant, and so a has at least one friend, b , and $n, b \models (Y \wedge FN)$. So a is in state N. Also a satisfies SY and so $n+1, a \models Y$, contradicting the stability of a 's preferences at u in M .

For the converse, suppose that $M, u \models \neg(\langle F \rangle \top \wedge \text{LOOP})$. Let a be an agent. Then either a has no friends, in which case her preferences are already stable, or a does not satisfy LOOP at u in M . We must show that a 's preferences at u in M stabilize.

Now consider the state machine depicted earlier and suppose for contradiction that that there is no $n \geq 0$ such that a has stable preferences in $\#_{\alpha,\beta}^n M$ at u . Since there are only a finite number of states, a 's preferences must loop. If so, for some n , a will be in either state Y or state N in $\#_{\alpha,\beta}^n M$ at u . This is because if a is initially in state Z then she cannot stay there, and Y or N are the only places to go. Also, by Lemma 5.2, it can never enter state I. So the only way to loop is by switching from Y to N and back. By symmetry again, we can assume that $n, a \models Y$ and $n+1, a \models N$. This can only happen by strong suggestion, so $n, a \models SN$, which implies $n, a \models FN$.

Consider a ' friends at stage $n+1$. We'll show that they can't all be in state Y. If they were, then for any friend b , $n, b \models SY$ because he has moved from N to Y and this can only happen if $n, b \models SY$. So $n, b \models FY$. And this holds for all a 's friends, and so $n, a \models FFY$. But we have showed that $n, a \models FN$ and so satisfies LOOP, contradicting our earlier claim.

So at least one friend b of a is in state I or N there being no other possibilities in the machine. Then $n+1, a \models ((I \vee N) \wedge \langle F \rangle (I \vee N))$, which is an invariant, by Lemma 5.1. So for all $m > n$, $m, a \models (I \vee N)$. But we showed above that a 's preference can never be I and so it is always N, contradicting our assumption that a 's preferences do not become stable. \square

6 Concluding Remarks

We have examined a particular model of the way in which preferences within a community can be affected by social relations. The model is highly idealised but still

displays an interesting dynamic structure, especially with regard to the moderating effect of indifference and the possibility of looping vacillations within the group. By using the language of ‘logic in the community’ we have showed that it is possible to characterize this behaviour with a simple formula. This is an initial test case of our methodology. Many variables can be changed to model different kinds of dynamic social situations. Among the possibility are the treatment of knowledge or belief instead of preference and the consideration of asymmetry social relationship such as subordination and aggregation.

The most pressing issue, however, is that of looking at transitivity-preserving preference upgrade operators. As we have shown, strong upgrade is an essentially non-deterministic operation, and so leads to non-deterministic dynamics, which should be interesting to investigate.

The application of modal logic to study social dynamics of this kind is new (insofar as we know).⁷ The usual approach is to use statistical modelling, typically accompanied by computer simulations. This research dates back to [3], which lays out a number of postulates concerning the degree of influence of one agent on another agent’s opinions, and proves various theorems about the resulting stabilisation patterns, specifically those concerned with whether or not a common opinion is formed. A recent example of research in this tradition is [4], which provides a good introduction to the mathematics involved and contains an extensive bibliography. The authors’ focus is the dynamics of ‘opinion fragmentation’ of which the formation of a consensus and the polarization of opinion are special cases.

Measured against this body of research, the contribution of the present paper can be seen as working along an orthogonal direction. Whereas most research in this area aims at giving more and more complex and accurate models of ‘influence’, we take a very simple model and show how to account for complex statements about it. As with all logic-based research, our emphasis is on what can be expressed in a particular formal language. A smaller, but significant, point is that our model is based on a very general model of preference, which includes the distinction between ‘indifference’ and ‘conflict’. Most work in this area takes a metrical approach, measuring opinion on a linear scale - or, more generally, as a vector in a larger dimensional space. Again these choices reflect a difference of focus. We are interested in the relationship between preferences between states and preferences between descriptions of states, and for this purpose the distinction between indifference and conflict is important.

Despite these differences, it is clear that existing research on social dynamics in psychology and AI contains many ideas that would contribute to a better understanding of the relationship between logic and society of the kind pursued here. In particular, the interest in opinion fragmentation could provide a focus for future

⁷ [2] is titled ‘A Logic-based Architecture for Opinion Dynamics’ but there is no formal language or semantics as such. The ‘logic’ used is fuzzy logic, and it is used only in the sense that opinions are modelled as fuzzy sets. This approach suggests connections to the psychological literature that may be interesting to explore further but there is little direct comparison to be made with the present paper.

research.

References

- [1] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
- [2] Piter Dykstra, Corinna Elsenbroich, Wander Jager, Gerard Renardel de Lavalette, and Rineke Verbrugge. A logic-based architecture for opinion dynamics. In *Proceedings of the World Congress of Social Simulation*, 2010.
- [3] John French. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.
- [4] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.
- [5] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, ILLC, University of Amsterdam, 2008.
- [6] Jeremy Seligman, Fenrong Liu, and Patrick Girard. Logic in the community. In Mohua Banerjee and Anil Seth, editors, *ICLA*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188. Springer, 2011.