Casilli, Antonio A., Rouchier, Juliette & Paola Tubaro (2014) How to build consensus in a health-oriented online community: modeling a «pro-ana» forum, *Revue Française de Sociologie*, vol. 55, n. 4, pp. 731-762.

How to Build Consensus in a Health-Oriented Online Community: Modeling a "Pro-Ana" Forum

Antonio A. CASILLI[1], Juliette ROUCHIER[2], Paola TUBARO[3]

[1] Telecom ParisTech / Centre Edgar-Morin—CNRS–EHESS, antonio.casilli@telecom-paristech.fr

[2] Groupement de Recherche en Économie Quantitative (GREQAM), Aix–Marseille Université, juliette.rouchier@univ-amu.fr

[3] Department of International Business and Economics, University of Greenwich, p.tubaro@greenwich.ac.uk

Abstract*. This article presents an agent-based model of a health-related Internet forum. If recent literature demonstrates the relevance of network approaches to gain insight into consensus-building within online groups of peers, the dynamic process of mutual adjustment of participants' health orientations has been seldom explored. Our model is informed by qualitative data collected via semi-structured interviews with Internet users living with eating dis orders—often stigmatized due to the controversies surrounding "pro-ana" (anorexia) websites. The discussion threads that unfold in the forum, expressing a range of health orientations from extreme "pro-pathology" to "pro-recovery" ones, initiate a mix of conflicting and supportive reactions that can trigger change in members' orientations over time. We develop a computer simulation of message exchanges i n a forum, describing micro-behaviors through a simple mechanism of influence. We then complexify the macro-setting, considering the effects of turnover (the possibility of exiting and/or entering the forum), and different rates of active participation of members to discussions. Our model shows that under empirically plausible conditions, moderate pro-recovery orientations are more likely to emerge than radical ones refusing medical mediation. These results lead to policy recommendations to design successful health information campaigns, and advocate against access restrictions or filtering of these online communities.*

Keywords. EATING DISORDERS, ONLINE HEALTH COMMUNITIES, E-HEALTH, SOCIAL INFLUENCE MODEL, AGENT-BASED COMPUTER SIMULATION

The present article is part of the research project "Ana-Mia Sociability: An Online/Offline Social Networks Approach to Eating Disorders" (ANAMIA), supported by the French Agence Nationale de la Recherche (grant ANR-09-ALIA-001). Legal and ethical approvals have been obtained from the French Commission Nationale de l'Informatique et des Libertés (CNIL), and the Research Ethics Committee of École des Hautes Études en Sciences Sociales (EHESS). The Netlogo model developed for this article is released under GNU GENERAL PUBLIC LICENSE (GPL v2) and is available online at: https://github.com/Bodyspacesociety/ANAMIA_F. We acknowledge help from Doctissimo.fr and B-eat , the British Eating Disorders Association, in advertising our survey to participants. We thank Lise Mounier, Fred Pailler and reviewers of Revue Française de Sociologie for valuable comments, and we are grateful to Sandrine Bubendorff for outstanding research assistance. The usual disclaimer applies.

Modeling conflict and support in online health groups

Conflict and support have been at the very center of sociology as a discipline, since the seminal contributions of its founding fathers (Durkheim 1893; Simmel 1908) who first discussed how human relationships are structured around discord and mutual help. Over the past few decades, web-based networked interactions have provided a unique field of observation and speculation, allowing social scientists to explore how these motives aggregate individual viewpoints and behaviors and shape the dynamics of participation and social influence among peers. Using agent-based modeling, in this article we set out to represent the role of conflict and support as structuring dimensions of the functioning of self-moderated communities.

To instantiate our analysis, we examine computer-mediated interactive forums centered on health and eating. The effect of peer-to-peer online groups on the formation of orientations and social norms about health and well-being is traditionally evaluated by analyzing the contents of pages and messages or—in moderated venues—via before-and-after comparisons (Eysenbach et al. 2004). But a more precise assessment of the efficacy of stand-alone interventions, or of the exactitude of medical information circulating on websites, discussion groups, and social media requires studying the structural social dynamics that enable adoption of specific health orientations within online communities.

Previous studies have demonstrated the usefulness of structural approaches based on social network analysis to provide insight into the formation of health orientations in self-styled online communities (Valente and Fosados 2006). However, there is still limited knowledge of situations in which diffusion of health behaviors does not result from top-down introduction of a healthcare inno- vation targeting a specific demographic segment, but rather from self-moderated interpersonal communication through online networks. In such contexts, orienta- tions emerge from individuals expressing conflictive or supportive interest in others' conducts or assertions. Agreement and disagreement are essential to compare and autonomously weigh alternatives in terms of health behaviors (e.g., treatment options, access to complementary therapies, etc.). Thus, the study of consensus building and conflict resolution processes needs to take into account the specificity of deliberative processes in this particular social space. As far as they do not simply derive from pre-established programs or explicit arguments in support of one course of action, but rather from "loose commitments" to ad-hoc therapeutic options or statements about health and illness (Akrich and Méadel 2007), orientations in these communities appear as the outcome of a collective negotiation aimed at reaching consensus via the conciliation of diverse, and poten - tially antagonistic, stances. As individuals share emotionally charged, and medi - cally relevant, personal narratives and digital contents (texts, images, links), they acknowledge the orientations of others and reconsider their own, in a process of mutual adjustment. Within this framework, the notion of "orientation" translates the object of study more accurately than that of "opinion," commonly found in much agent-based modeling literature. If the latter notion focuses on judgments of circumstances and facts with respect to their character, merit, or features, the former brings forward more subjective and emotionally-charged stances.

Internet communities centered on eating disorders are the empirical context of this article. Often negatively labeled "pro-ana" (anorexia nervosa) and "pro-mia" (bulimia) and conflated with a "social movement" advocating and glorifying eating disorders (Casilli 2013), this particular segment of the international web has come to the fore of public debates for hosting tips on starving and purging together with doctored photos of celebrities supposed to inspire thinness (hence dubbed "thinspiration"), and the disturbing, albeit infrequent claim that eating disorders are a lifestyle choice rather than a disease. Yet, these communities also provide alternative tools of self-help and support for sufferers and sometimes accompany them toward treatment and recovery. This ambivalent and paradoxical posture (Yeshua-Katz and Martins 2013) makes it difficult to predict the effects of participation on members' health orientations and—ultimately—on their eating behaviors and health state. The most appropriate policy response is also unclear. Reactions so far—stigmatization, ban from many web services, and threats of restrictive legislation (Casilli, Pailler and Tubaro 2013)—have been based on thin evidence and may prove to be counter-productive.

To assess how participation in such a forum shapes personal health orienta - tions, we combine use of qualitative interviews[i], providing insight into modes of discussion, degree of personal engagement, and duration of membership; and an agent-based model, ANorexiA-buliMIA Forum ("ANAMIA_F"), reproducing these essential features in silico to observe the global patterns that arise from them, as well as to compare and contrast the effects of different underlying condi- tions, factors and contexts[ii].

We build on the wide agent-based simulation literature that since the 1990s has extensively investigated social influence and its population-level effects follow- ing, among others, Serge Galam and Serge Moscovici (1991), Robert Axelrod (1997), Guillaume Deffuant et al. (2002), Rainer Hegselmann and Ulrich Krause (2002). The agent-based approach has the advantage of allowing quasi-experi- mental research conditions, insofar as simulators create an artificial society of which they control all parameters. By varying them ceteris paribus and simulating the behavior of the system under the changed conditions, researchers can observe the ensuing differences in outcomes—notably in this study, the extent to which forum members become more radical in their opposition to medical and professional healthcare mediation ("pro-pathology" stance), or more open to collaboration and treatment ("pro-recovery"). It is possible in this way to assess how the emergence of pro-pathology or pro-recovery orientations depends on rules of individual behavior, such as members' initial individual orientations and openness to influence, and on structural factors, such as the visibility of the forum and its accessibility to new members.

Qualitatively informed agent-based simulation

Although agent-based models are particularly valuable to investigate network dynamics and influence (Rouchier, Tubaro and Emery 2014), they are more gene - rally used in research on complex social systems and adaptive behaviors among heterogenous interacting individuals (Gilbert and Troitzsch 1999); they have also been employed to study computer-mediated interactions (Schweitzer and Garcia 2007; Ren and Kraut 2014), social media and other sharing platforms (Zhou et al. 2011).

The building blocks for the design of an agent-based model are the definition of a population of artificial agents, the establishment of a set of interaction rules and the assessment of the effects of these interactions on the agents' state and position in a given behavioral space. The way these building blocks are created, differentiates agent-based models into two main families: abstract and empirically- informed simulations. The former are built by abstraction from a target system (a social phenomenon) and are meant as a guide for theoretical reflection, the latter use qualitative and/or quantitative data collected in a given social context to define simulation rules and parameters, so that the interpretation of results is in closer link with a precise field-based question.

Our model falls into the second category. We adopt what can be described as qualitatively-informed computational approach. Although the use of quantitative data has been dominant so far among empirical agent-based modelers, qualitative data have been employed for validation, rule-definition and parameter assessment since the late 1990s (Chattoe 2002). This method builds on the idea that, despite their formal nature, "there is nothing inherently quantitative" in agent-based simulations (Yang and Gilbert 2008: 175). Conversely, there is growing awareness among qualitative researchers that "the world of agent-based modeling is ethnographically seductive" (Agar 2003). A general framework to develop qualitatively- informed agent-based modeling is presented in P. Tubaro and A. Casilli (2010) and summarized in Figure 1.

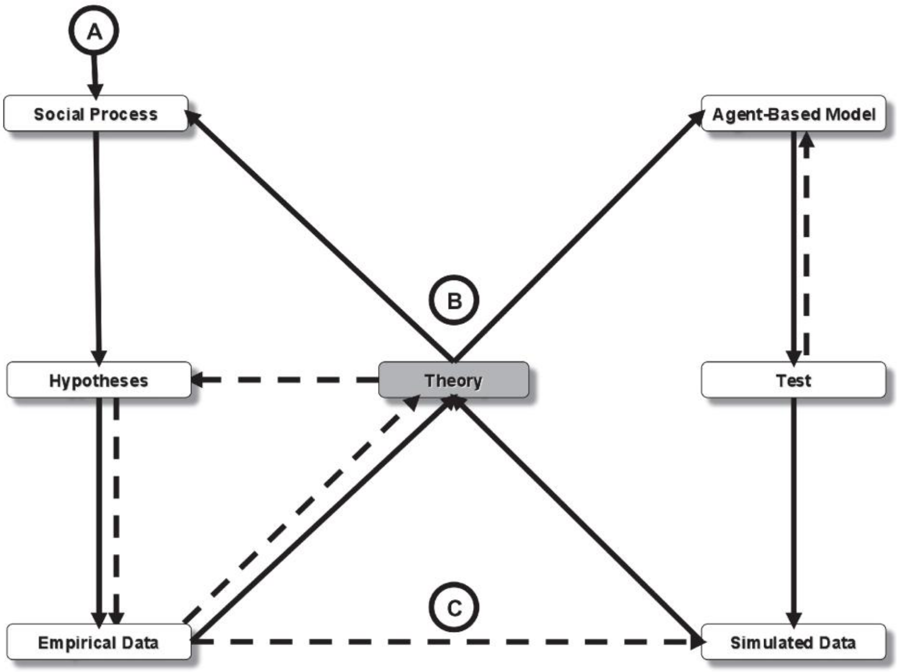

Figure 1 - Design procedure for a qualitatively-informed agent-based model

The starting point (A) is a target social process representing the object of an empirical study. After research hypotheses are formulated and empirical evidence collected, theoretical elements can emerge (B). Subsequently, an agent-based model is designed as a proof of concept, tested and corrected. The model produces simulated data that can be compared to empirical data (C), allowing to reformulate the theory. Additional steps for empirical validation can be added (dashed lines). Source: Tubaro and Casilli (2010).

The resulting model is based on hypotheses that are instantiated by qualitative evidence (Amblard, Bommel and Rouchier 2007). When simulations are run, parameters and behavioral choices shape a succession of time-steps during which agents evolve individually and adapt mutually. The iteration of these micro-level behaviors can lead to meso- or macro-level regularities, which emerge over time (Manzo 2007a). Analysis of the simulated data aims to reveal regularities that are not expected ex ante, since they are not inbuilt in the model, but rather result from iterative interactions. Usually, these behavioral patterns appear for certain parameters only (Grimm et al. 2006). After ascertaining the linkages between patterns and parameters, a comparison with empirical data allows to loop back on the "target system" so as to provide better insights and refine research questions[iii]. Along these lines, we now summarize essential insight from the fieldwork and then outline our use of it to inform the ANAMIA_F model.

Salient facts from the ANAMIA fieldwork about eating disorder forums

As part of the ANAMIA research project, we have undertaken a three-year long study of Internet use and health orientations of members of the web community of persons living with eating disorders in France and United-Kingdom. The study consisted initially in the crawling and analysis of a corpus of English- and French-language eating disorder blogs, forums, and personal web pages over the years 2010–12. Subsequently website users (N = 284) accepted to respond to an online questionnaire about their medical history, Internet use, personal network (both face-to-face and computer-mediated), and socio-economic status. Finally, a subset of this population agreed to an in-depth interview, lasting from 50 to 90 minutes, conducted by phone or Skype (N = 37) (see Table 1 for sample characteristics). All the interviews were transcribed verbatim and exploratory text analysis was performed using qualitative data processing software (Alceste). Afterward, the interviews were read iteratively by the authors and other ANAMIA team members in order to bring forth discursive trends, and to detect specific categories of users and typical use scenarios.

		FR	UK
N		22	15
Gender		100% F	100% F
Age	*Mean*	23	23
	Min	18	16
	Max	38	37
N students		13	11
N in paid work		4	7
N with own children		1	1
Body-Mass Index[1]	*Mean*	21.4	18.1
	N underweight	7	12
	N overweight	4	2
Eating disorders	*N anorexia nervosa*	7	11
	N bulimia nervosa	10	0
	N EDNOS[2]	4	4
	N binge eating	1	0
Treatment	*N treated*	10	12
Time online per week	*N low (< 14 hours)*	8	8
	N medium (15 - 24 hours)	4	3
	N high (> 25 hours)	10	4

Table1 - Sample characteristics, absolute figures, for France and United-Kingdom interviews

[1] The body-mass index (BMI) is a measure for human body shape based on an individual's weight and height. The World Health Organization uses it to distinguish underweight (BMI < 18.5), normal weight (18.5 < BMI < 24.9), and overweight (BMI > 25) in adults.
[2] EDNOS = Eating Disorders Not Otherwise Specified, primarily mixed and attenuated forms, as defined in the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (DSM IV).

The resulting themes and lexical classes were used to extract the following stylized facts.

1) Conflictive and supportivemessages shape up the dialectical interaction of pro-pathology and pro-recovery attitudes

Interactions in eating disorders-forums consist of messages whose contents fall in a continuous spectrum going from pro-pathology to pro-recovery extremes. The general response to any single message can be supportive (agreement) or conflictive (disagreement). The overall orientation of the forum will result from the aggregate composition of all messages.

In this sense, no forum is per se a pro-ana community, yet it can occasionally manifest a more clearly pro-pathology tone. The general situation is more reminiscent of a dual orientation, where acceptance and refusal of medical mediation coexist.

The goal here was not to recover or to go further into anorexia. It was just to find people who lived the same thing as me. Just that. (Excerpt interviewee ID 78 FR).

Interviewees display a specific tendency to fluctuate between the two. One respondent goes as far as to describe that as an agony of indecision.

It is a bit of a tension because ... you try to get out of it somewhat, but at the same time, you don't want to abandon the people there. (Excerpt interviewee ID 13 FR).

Interviewees are ambivalent about their own perception of the so-called "pro-ana" phenomenon. Their first motivation to contribute to eating disorder forums is the quest for help and solidarity—a "sense of community."

I certainly get a lot of solidarity and support I think. … barriers are often down and they're often less inhibited and more… more open to erm… revealing personal information and... and sharing their own experience and that... really create sense of community and.... it just feel that you're not alone. (Excerpt interviewee ID 610 EN).

Only a tiny minority expresses explicit pro-pathology views, although a large majority is aware of them. In fact, interviewees tend to distance themselves from the "pro-ana movement" in order to forgo the stigmatizing label it bears. To some extent, there is also a process of internalization of the biomedical norm. The "real pro-ana" (that is, more radical pro-pathology postures) appear as harmful and dangerous, or as an obstacle in the way of recovery.

It is true that when I'm not well, I often go on these sites and I know I should not because... on the contrary, they should be shut down, it's ... well, it's disgusting. (Excerpt interviewee ID 83 FR).

However, anti-recovery stances can kick in at any moment, without discontinuity. Pro-ana attitudes may offer an alternative in those cases when recovery seems impossible. It is at this point that sometimes a pro-ana orientation can be regarded as a discursive wave that members can exploit to reach some form of online popularity or social support (agreement). It becomes a rewarding way to manage the distress, and to overcome the social exclusion, that are often associated with eating disorders.

When you didn't follow, you became the black sheep. And that's how, in the end, the girls end up in the so-called pro-ana groups... because, when you can't have more on the forums ... if you cannot talk about what's wrong, and you get reproached because you can't get better, well, that's hard, you know. (Excerpt interviewee ID 11 FR).

Yet pro-ana is never uncritically accepted. Contributors posting explicit pro-pathology messages can become the target of manifest hostility and their contents are discussed, criticized, edited.

Well, there was one who said things that uh... upset us all, it was pretty horrible what she said and uh... and we did not really want that to happen again, in fact. She said that... people uh... that obese people do not deserve to live, that... that uh... pro-ana that was the only thing that mattered, uh... there was nothing better than feeling thin, etc. ... Her posts have been deleted because it... it was a little sad, actually... to keep them on the forum. (Excerpt interviewee ID 9 FR).

2) Eating disorder web forums are characterized by high turnover and frequent member entries and exits Forums are highly volatile social structures.

Given their voluntary nature, their existence can be disrupted or discontinued at any moment. More importantly, they constantly transfer to new servers, new platforms, new countries. This is due not only to technical constraints or personal circumstances of the participants, but also to legal reasons. Mass migrations of forum members from one community to another rely on informal mechanisms such as word of mouth.

> *I was on another forum, uh... that uh... was also closed, but by the administrator of the forum, uh... I don't know why. So, all the people who were on that one, we then gathered elsewhere, uh... in a very controlled forum where uh... if you do not register you do not have access to all categories, you must first fill out a questionnaire, etc., to know the person a little bit. (Excerpt interviewee ID 31 FR).*

For similar reasons, within a single forum, audience can be rather erratic. Eating disorder discussion websites display high turnover, and members are aware of the intrinsic instability of their online contacts.

> *Some connections developed, some connections died and, you know, life goes on… (Excerpt interviewee ID 607 EN).*

Entries are motivated by the need to get in touch with like-minded persons. Respondents try to find others living with eating disorders who share a common point of view, who experience similar situations, and process them in a way that resonates with their sensibility. The expression of agreement and support is crucial to allow comparison with others and self-evaluation, as well as to assess the fit of single members with the perceived overall forum orientation.

> *I'd say it's when you don't want to get better anymore and at the same time... need to feel you are not the only one, to remember uh... because when you read the other, you see yourself a bit like in a mirror. (Excerpt interviewee ID 12 FR).*

Exit decisions depend on the perceived prevalent and stable orientation of the forum. When members feel that others are progressively becoming more radicalized in a sense that is opposite to their position, and if this situation is perceived as settled (e.g., if messages are perceived to be too pro-recovery or too pro-pathology over a period of time) they can decide to quit the forum. This can be an individual move (departure) or a joint move (schism).

> *A forum like that other one, the crazy one, the dangerous thing, then... uh... but in this one, there was already a selection. This is the forum where I really... well, I was going all the time, and the discussion was uh... relatively proper. Uh... and uh... then, one of the moderators of this forum have created the migration forum. And uh... most people who were active on that forum migrated to the new one. And then it's... Ah, I know uh... this or that person, well, it's good, we accept you' and so on. (Excerpt interviewee ID 31 FR).*

3) Styles of web use and degree of participation to online forums differ

The degree and modes of participation in forum discussions vary widely. Some members contribute by actively posting contents.

> *On the forum it's also like... it's like going on stage in a sense... you show yourself... I talk to myself, it's a bit like keeping a diary, uh... but with readers. ... Readers and answers. Answers that are sometimes relevant and... ideas, impulses... uh... questions that direct me to new paths uh... and... So that too is enriching. (Excerpt interviewee ID 65 FR).*

Some members, instead, only react to other members' messages and post comments, but do not take the initiative to propose original content.

> *I somewhat fear of falling into the vicious circle, the circle "I posted a comment," then they will reply, and I have to re-reply … (Excerpt interviewee ID 83 FR).*

Finally, some members limit themselves to lurking, i.e., to passive participation.

I don't do anything there, it's... it's mostly to see the same people uh... uh... I go on the forums, but uh... in fact I speak with nobody anyway, I remain uh... I am invisible, a little ghost, whatever. Well... I... uh... I read... watch what others say, but uh... I never speak in fact. I do not have an account on the forums uh... (Excerpt interviewee ID 131 FR).

These three behaviors (posting, commenting, and lurking) are customary in most self-moderated online platforms; the interesting thing here is that by and large, they correspond to different phases in the online trajectory of members of the online community of persons living with eating disorders.

Lurking can be equated to exploratory information collection, one of the reasons why persons with eating disorders start exploring dedicated websites and forums at the beginning. Efforts to collect, evaluate and use online health information (about types of symptoms, existence of communities of patients with the same disorders, recommendations for treatments, etc.) are most common at the stage of onset, or initial awareness, of an eating disorder.

It's a bit... voyeuristic. You could see how people are doing and... compare to them. (Excerpt interviewee ID 643 EN).

A second stage involves reading, without directly engaging in online conversations. Reading blogs and forums, even without interacting with other participants, may contribute to breaking the sense of isolation that the disease often brings about.

I generally look at... at what... the people are saying, rather than write anything or reply or anything like that. I'm just a stalker I guess. (Excerpt interviewee ID 640 EN).

A member who does not post written messages is not necessarily excluded from the interaction. Aggregation of orientations can still be achieved by comparing "stats," i.e., personal statistics and automated metrics.

I think... "Halls MD"... I don't know, I have it in the bookmarks... and that has like BMI calculator that puts in your weight-loss percentile, like, what category in... how many people have bigger percentage. (Excerpt interviewee ID 607 EN).

Research questions

In light of the empirical elements presented above, we pose the following research questions for our ANAMIA_F agent-based model:

- RQ1: (cf. stylized fact 1). Under which conditions does ambivalence leave way to a shared orientation within an online forum?
- RQ2: (cf. stylized fact 2). Does easiness of entry and exit in a forum encourage the emergence of a shared orientation?
- RQ3: (cf. stylized fact 3). How does the changing ratio of active participants over passive lurkers affect the forum's overall orientation?

The last two research questions refer to structural factors that can be, at least partly, controlled by the policymaker—for example, free access to forums rather than filtering or censorship may affect the rate of turnover or the ratio of active vs. passive members. Assessing through simulation the likely effects of these factors, and thence of possible health interventions, is a way to provide guidance for health policy before any actions are actually undertaken.

To investigate our research questions, we advance the following hypotheses:

- H1: If no exit is allowed and there is dominance of negative commitment (disagreeing and conflictive answers to posts), no consensus can be reached, as mutual influence among forum members does not find a middle ground between diverse stances (whether pro-pathology or pro-recovery). This results in a polarization scenario, with coexistence of two orientations remaining very far apart from each other, each aiming unsuccessfully to become a norm.
- H2a: If exit is allowed and there is limited acceptance of diverging stances, one orientation prevails, due to the exclusion of disagreeing agents. This results in a radicalization scenario (one orientation becoming the norm for all members).
- H2b: If entry is allowed and there is a limited acceptance of diverging stances (whether pro-pathology or pro-recovery), separate groups will form and a polarization scenario will emerge.
- H3: If the number of participants is proportionally higher relative to the number of lurkers, a polarization scenario is more likely to emerge.

Key features of the ANAMIA_F agent-based model

Building upon our empirical data and the existing agent-based literature, the ANAMIA_F model features: 1) an influence mechanism, typifying micro orientation adjustments for individual agents facing a discussion thread, 2) an interaction structure describing the discussion thread in reaction to an original message, 3) a framework of participation representing the macro structure allowing to entry, contribute to, and exit from, a web forum. We outline here the essential elements of each of these components and give a formal, more detailed description of the model with the complete list of variables and parameters in Appendixes 1 and 2.

1) At the heart of our model lies the assumption that each agent is defined by an orientation, and this orientation can be influenced by the others. From among existing agent-based representations of influence, we borrow the one of Wander Jager and Frédéric Amblard (2005) which best embodies the insight from fieldwork that agents can feel positive, negative or neutral regarding the orientations expressed by the others. This approach features continuous representations of agents' inner states, taking a real value in the interval [- 1; 1], though we interpret them as orientations rather than opinions. The choice of using continuous values is now popular (Deffuant et al. 2002; Hegselmann and Krause 2002; Kozma and Barrat 2008; Rouchier and Tubaro 2011) because it accounts for a range of positions between two extremes, and it is thus preferable to binary approaches (Galam and Moscovici 1991); although it represents only one topic or one dimension of the orientation, it avoids the complexities that are known to arise with multiple dimensions (Rouchier and Tanimura 2012).

Most agent-based representations of influence include a principle of boundedconfidence (Hegselmann and Krause 2002): agents will influence each other only if their views are close enough. In a continuous context, this means that each agent has a range of perception (called "latitude of acceptance," lat^a in [0; 1] in our model) defining an interval within which other agents' orientations can influence it positively. This positive influence is modeled as a translation of the agent's orientation in the direction of the others'. What W. Jager and F. Amblard (2005) add is the idea that agents can also influence each other negatively: if someone expresses a truly disturbing orientation, the agent will react by widening the gap between its own orientation and the other agent's. Thus, a second interval of perception must be defined, namely a "latitude of rejection" (lat^r in [lat^a; 1]). In this framework, whenever an orientation is expressed, it can diminish the gap between two orientations (positive), widen it (negative) or keep it unchanged. Therefore, lat^a and lat^r can be taken to represent different levels of agent permissiveness, narrow-mindedness and neutrality.

Our simulations test the whole spectrum of values of these two parameters, whose combinations represent the wide variety of human attitudes, cannot be narrowed down on the basis of empirical evidence, and are hardly affected by policies.

2) The interaction structure is a completely original element of our design, and suits the context of online forums on eating disorders. The interaction takes the shape of a topic thread in a conversation. One randomly chosen agent sends an initial message about eating disorders (personal experience, health information request, sentiment expression or preference revelation). Any other agent can reply, provided it perceives the content of the original message. Specifically, if the content falls within its acceptance (respectively rejection) intervals, it will answer with a supportive (respectively conflictive) message containing its own orientation (Figure 2). Otherwise it will not react.

An indicator of the intensity of interactions is the number of supportive and conflictive replies that each original message elicits. While most agent-based literature considers dyadic interactions, we deviate from this approach and, to the purpose of closely mimicking the actual dynamics of conversation and influence, we assume that messages can potentially affect all forum members: the original sender, all replying agents, and those who do not actively participate, depending only on their latitudes of acceptance and rejection. All supportive messages are averaged in one supportive reply, all conflictive messages in one conflictive reply. Both these averages can influence all agents, and all update their orientations accordingly. This collective effect mirrors insight from the fieldwork that comparison between individual stances at dyadic level (agents i and j) is less relevant than affinity between a single orientation (agent i) and the general ambiance of the forum (agents j, k, l, etc.).

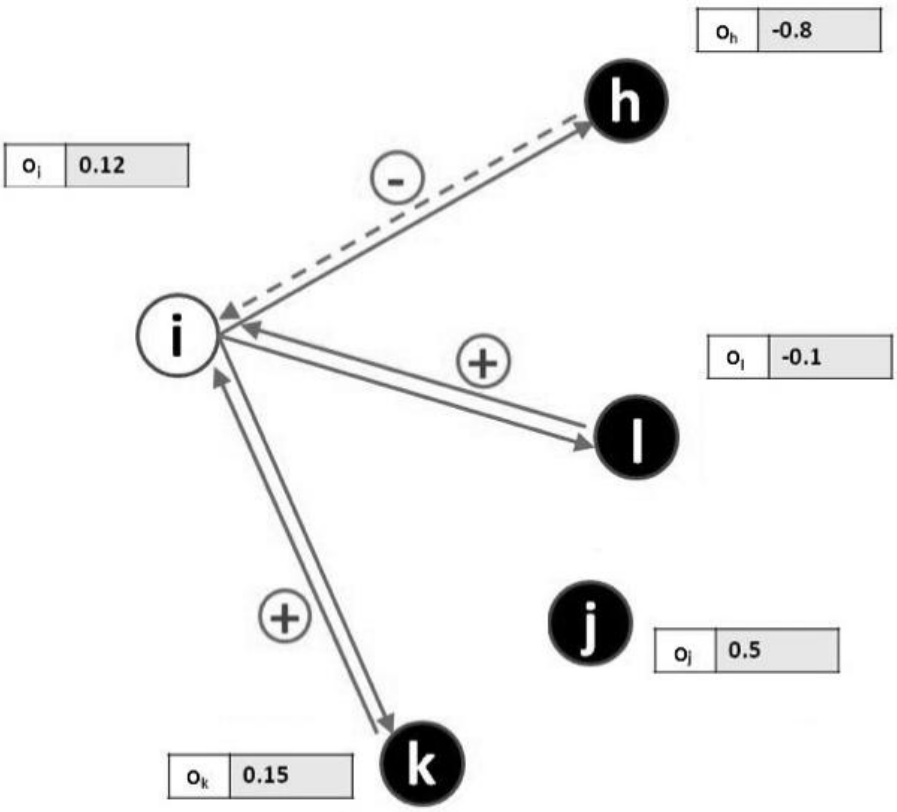

Figure 2 - Example of a message and the replies it elicits in our model

Each agent is characterized by a value of o (orientation). Agent i starts a topic thread sending an original message; agent h engages a conflictive reply (dashed line); agents k and l express support (solid line); agent j is neutral (does not perceive the original message and does not reply).

The interaction structure shapes the encounters that give rise to influence and can be seen as the most original part of our model. Very few agent-based simulators have modeled individual-group interactions and the influence that may derive from them. Either, they consider that group communication is globally based on the repetition of numerous dyadic interactions—following G. Deffuant et al. (2002)—or that all agents communicate with others within their confidence interval, but without asymmetry—following R. Hegselmann and U. Krause (2002). Another exception is Dirk Van Rooy (2012) who attempts to match simulated and experimental data, but only for small-group communication.

3) The framework of participation is also a novelty of our model, informed by the fieldwork. Firstly, we summarize in one parameter (rate of participation) the unevenness in interactions stressed by our interviewees, and also observed in other, non-eating disorder-related forums (Hargittai and Walejko 2008). Existing literature suggests that user-generated contents may follow an inverse power law, epitomized by the operational rule widely known as the "90–9–1 principle," defining gross proportions between different roles and participation styles. Although empirical estimates of effective participation in online discussion are still tentative (Ochoa and Duval 2008), the rule calls attention to the fact that a small minority of content creators (the quintessential "1%" of members posting messages to the forum) coexists with a larger group of sharers and commentators (the so-called "9%") as well as a vast majority of passive readers (those "90%" viewing, but not

replying to, online messages) (Nielsen 2006). This element is integrated into our model so that, for each new thread, only some of the forum members will answer. The rate of participation represents their proportion. We initially fix this parameter at 10 % (close to the "9%" of the literature), and then vary it to test its specific effects. For the sake of generality, we do not introduce any exogenous assumptions about which agents are more likely to initiate threads, and presume that they are equally likely to do so.

Secondly, the framework of participation defines the dynamics of turnover (exits and entries) in forums. We test three versions of the model: 1) no turnover, 2) exit is allowed when agents are too dissatisfied of other members' reactions to their messages, and 3) both exit and entry of new agents are allowed.

A situation in which exit is allowed, but leavers are not replaced, mimics policies where eating disorders forums, perceived as dangerous by web service providers or policymakers, endure restrictions to entry (e.g., imposed warnings and disclaimers, website filtering, or search engine censorship as discussed later), to prevent them from advocating radical "pro-ana" stances. Exit depends on a patience parameter. When an agent launches a thread, it counts the occurrences in which conflictive replies outnumber supportive ones; if this happens over a succession of time-steps higher than patience, then this agent (and only this agent) quits the system.

When entry is allowed, we consider that the replenishment of the community takes place as members leave, and we test two versions. In the first, incoming agents are allowed in to replace exiting ones, regardless of their initial orientation, which may be different from the general ambiance and consequently introduce conflicting stances in the community. In the second version ("selective entry"), only agents of a particular type, notably with a "positive" orientation (in [0; 1]) enter the community. Theoretically, this procedure builds on previous research that reveals the important effects of turnover and social influence, on systemic outcomes (Rouchier and Tubaro 2011; Rouchier, Tubaro and Emery 2014), and that points to newcomers' traits as factors that shape the outcomes of the simulations. In policy perspective, selective entry stands for the introduction of "ambassadors" of recovery-oriented stances, who may be sent by charities or public health bodies to direct persons with eating disorders to treatment.

Experiments, parameters, and indicators

Table 2 summarizes the values of the main parameters included in the model. Most interesting to answer our research questions are lat^a and lat^r representing, respectively, the scope for positive and negative influence; the rate of participation, which reflects the different possible proportions between posters, commentators and lurkers; and exits and entries, which account for turnover.

Parameter	Value
Size of population	200
Influence μ	0.1
Patience p	2
Latitude of rejection lat^r	$[lat^a; 2[$
Latitude of acceptance lat^a	$[0; 2[$
Rate of participation	*10, 40, 70, 100*
Exit	*on, off*
Entry	*on, off*

Table 2 - List of main model parameters with the values used in the simulations

We do not vary the parameters that do not directly refer to our hypotheses, and use the reference values of W. Jager and F. Amblard (2005). The full list of parameters and notations is in Appendix 2. Reading: List of main model parameters with the values used in the simulations; those being tested are in italics.

We observe indicators of the evolution of orientations at aggregate level, looking at mean values above 0 ($avg(o_{pos})$) and below 0 ($avg(o_{neg})$), and the related standard deviations, when the system has reached a stationary state. Indeed, we aim to establish whether, and under which conditions, the forum splits into sub-groups with opposed orientations, or reaches an agreement around a commonly shared view. We also aim to assess in which conditions extreme rather than moderate views prevail. To disentangle the effects of influence and turnover, we also observe the size of the population at the end of a simulation when exit is possible, and the rate of turnover when agents can enter as well.

Results of the simulations

Four different outcomes can emerge

The outcomes of the simulation vary according to agents' characteristics (primarily parameters lat^a and lat^r), and structural factors, namely turnover and rate of participation. In all settings, agents initially exchange messages and their orientations evolve until stabilizing—with no more individual changes and a constant average value. From then on, if exit is authorized, unsatisfied agents leave. The fact that agents do not do so before this stage is not built in the model, but depends on the speed of convergence which in all our simulations, is higher than the patience of agents (both constant). Entry of newcomers to replace leavers, when allowed, destabilizes mean orientations and re-launches the process, which continues until mean orientations and the size of the population stabilize again. Four outcomes can be observed at the end (Figure 3):

1) *Forking Polarization*: The community is split into two groups whose orientations lie at the extremes of the admissible range, - 1 and + 1; the two groups need not be of equal size. For operational purposes, we consider that this result is obtained when $0.9 \leq (avg(o_{pos})) \leq 1$ or $- 1 \leq (avg(o_{neg})) \leq - 0.9$.

2) *Parallel Polarization*: The community is split into two groups whose orientations lie closer to the midpoint of the acceptable range. These groups are not necessarily of equal size, and the values ($avg(o_{pos})$) and ($avg(o_{neg})$) do not necessarily have the same absolute value. For operational purposes, we consider that this result is obtained when ($avg(o_{pos})$) < 0.9 and ($avg(o_{neg})$) > - 0.9.
3) *Winner-Take-All*: Only one orientation is shared among all members of the community; it may be positive or negative and it is beyond the threshold for consensus (that is, ($avg(o_{pos})$) ≥ 0.1 or ($avg(o_{neg})$) ≤ - 0.1), indicating that the community has taken a decisive drift in one direction.
4) *Consensus*: Only one orientation is shared among all members of the community; it may be positive or negative, and is close to the midpoint of the acceptable range (0) so that it can be considered close to neutrality. We operationalize this notion by considering that there is consensus when ($avg(o_{pos})$) < 0.1 or ($avg(o_{neg})$) > - 0.1.

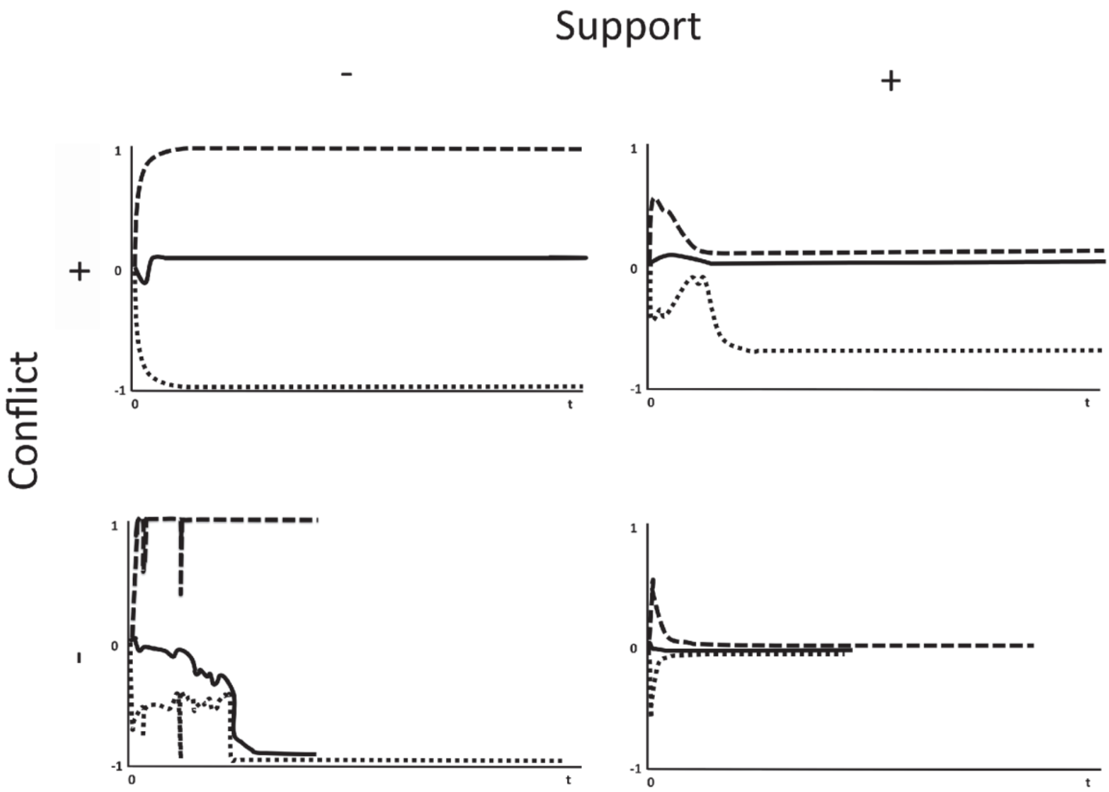

Figure 3 - Examples of the four possible outcomes

Upper panel, left: Forking Polarization, right: Parallel Polarization. Lower panel, left: Winner- Take-All; right: Consensus. Time is on the horizontal axis, average orientations on the vertical axis.Dashed line: Average of positive orientations, Dotted line: Average of negative orientations, Solid line = Average orientations for the whole community.

In the two polarization cases, the community is split into two sub-groups (not necessarily of equal size) sharing different orientations, while in the other two cases the community gathers around one single orientation. Both types of polarization describe conflictual situations, where the path to equilibrium is paved with disagreement and personal messages are constantly challenged by others. The cases of Parallel Polarization and Consensus reflect a more supportive community in which prevailing opinions tend to be more moderate, and a stationary state is reached when all messages and replies start to convey cooperation and solidarity.

Before exploring the conditions under which these different scenarios emerge, it must be mentioned that occasionally the system reaches a configuration of Forking Polarization but not a stationary state: agents cluster into two groups that share strongly differentiated orientations, at a distance from each other that disallows any further reciprocal influence, but there is significant variation within each group, and averages oscillate continuously around values situated close to the extremes. These fluctuations are within finite bands and the number of agents in each sub-group remains constant, suggesting that the resulting configuration is unlikely to evolve over time—similar to the punctuated equilibrium described in Joshua M. Epstein (2002).

Effects of individual characteristics: Latitude of acceptance and rejection

The main individual characteristics whose effects we explore are latitude of rejection and latitude of acceptance, lat^r and lat^a. We first observe their effects in baseline simulations with low rate of participation (10%) and no turnover. While the former structural criterion is a proxy for the "90–9–1" rule outlined above, the latter should be understood as a purely theoretical benchmark, a point of reference to be used for comparison purposes (see below), rather than an accurate representation of an empirically observed occurrence.

Forking Polarization ensues in all cases in which lat^r is low—in practice, this means equal to, or lower than, 1.1 regardless of the values of other parameters. Some of these cases are punctuated equilibriums. This result is highly stable: in particular, low levels of lat^r always produce a result situated in the left panel of Figure 3, corresponding to dominance of extreme orientations.

If lat^r is higher than 1.2, all four final outcomes can be observed. The fact that one of these possible outcomes eventually emerges depends on the value of the latitude of acceptance lat^a (see below). In baseline simulations, Forking Polarization and Consensus appear about 40% of the time, and Parallel Polarization appears 11% of the time (Figure 5). The latter two are cases in which more moderate orientations prevail, and convergence is achieved immediately in the first few time steps of a simulation run.

Latitude of acceptance lat^a determines the extent to which agents' orientations get closer to each other. While as discussed above, lat^a plays no role for $lat^r < 1.2$, the final outcome depends on lat^a for higher levels of lat^r (between 1.2 and 2). In general, higher levels of both lat^r and lat^a correspond to more moderate final orientations, whether in the form of Consensus or Parallel Polarization; in contrast, lower levels of lat^r and lat^a correspond to more extreme orientations, whether in the form of Forking Polarization or Winner-Take-All. Figure 4 represents average orientations for different levels of lat^a and lat^r in baseline simulations.

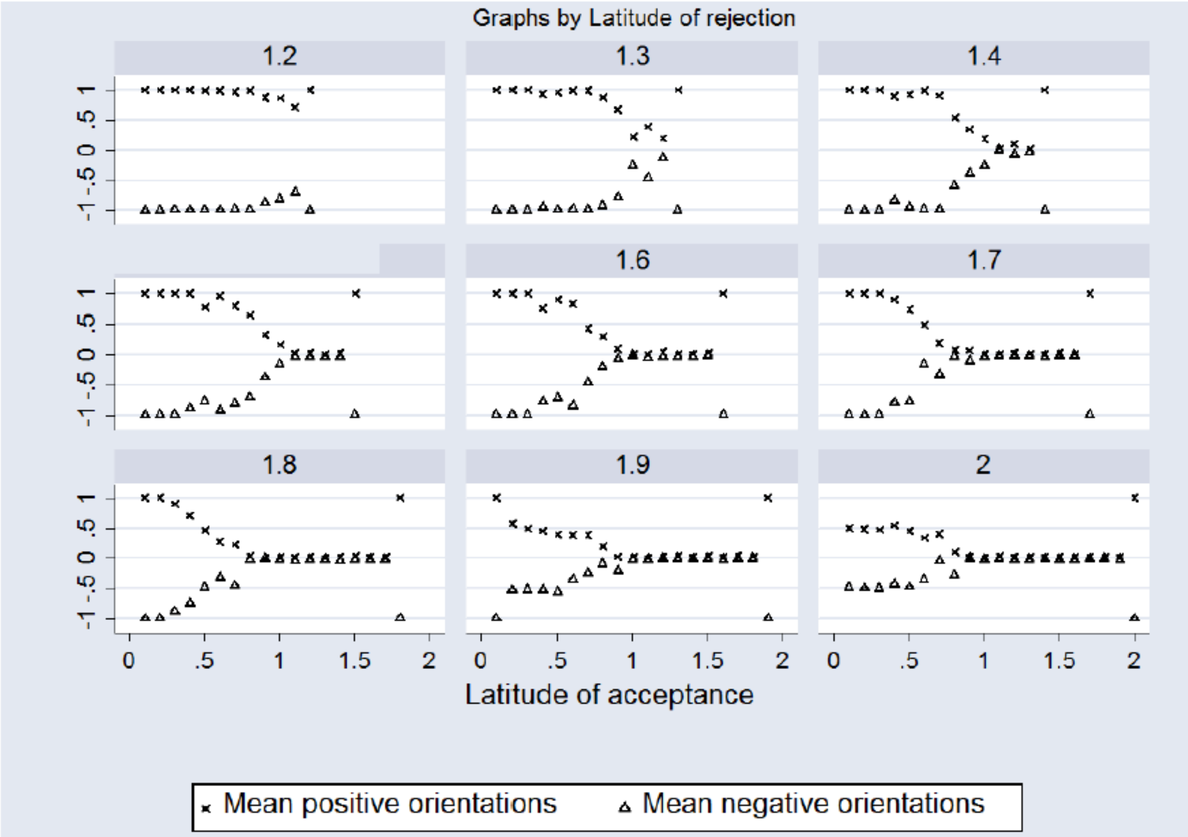

Figure 4 - Mean positive and negative orientations Graphs by latitude of rejection

The data used to produce the figure are averages taken over ten simulation runs for each combination of the parameters. Mean positive and negative orientations at the end of baseline simulation runs (y-axis), plotted as functions of the latitude of acceptance (x-axis), for different levels of the latitude of rejection in its higher range (1.2 to 2). For example, the bottom right panel shows that when lat^r = 2, the most likely scenario to appear is Parallel Polarization with convergence to orientations of +/- 0.5 for low lat^a, while all orientations tend to 0 (consensus) when $lat^a$is high.

Effects of structural factors: Turnover

Let us now remove the unrealistic assumption of no turnover and assume, first, that exit (but not entry) is allowed; rate of participation remains at 10% and we vary lat^r and lat^a in their range. Under these conditions, dissenting agents initially cluster around one extreme, in opposition to other agents grouped around the other extreme; then they leave the community, resulting in a Winner-Take-All situation (Figure 5, upper right panel). In fact by allowing one dissenting group to abandon the forum, exit transforms a polarized situation into a Winner-Take-All with final convergence to either + 1 or - 1 (both outcomes being equally likely to occur). Winner-Take-All always emerges when lat^r is below 1.1, and emerges about 40% of the time when lat^r is high—just the same proportions that Forking Polarization takes with no turnover (see above). The relative frequency of occurrence of Consensus and Parallel Polarization are not affected by the possibility to exit. During each simulation run, the population shrinks and falls to less than 10% of its initial size when lat^r is low, and less than 30% when lat^r is high. Only when

Consensus or Parallel Polarization emerges, does the size of the population remain about stable at its initial level. When newcomers are allowed to replace leavers, Forking Polarization becomes the most common outcome again, occurring all the time for low lat^r and 40% of the time for high lat^r (Figure 5). The frequency of occurrence of Consensus and Parallel Polarization remains unchanged. Turnover is low in the latter two cases, and high when Forking Polarization emerges.

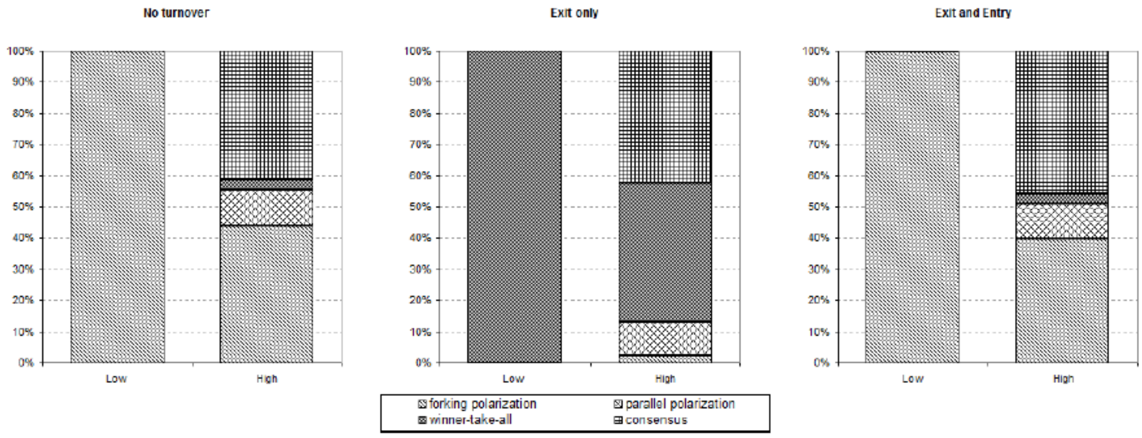

Figure 5 - Occurrence of each of the possible final outcomes in percentage over the whole set of observations

Results are based on ten simulation runs for each case, for all admissible values of the parameters. Occurrence of each of the possible final outcomes, in percentage over the whole set of observations, for low (< 1.2) and high (1.2 – 2) values of the latitude of rejection, when there is neither exit nor entry (top left), exit only (top right), both exit and entry (bottom). 100% of all simulation outcomes with low lat^r, and about40%with high lat^r, are Forking Polarization when there is no turnover, or full turnover with exit and entry; they are Winner-Take-All otherwise. Consensus and Parallel Polarization appear in about the same proportions under all circumstances, when lat^r is high.

A variation of the full-turnover model involves "selective entry," where only agents with orientation above 0 (which we interpret as pro-recovery stances) are allowed in. Table 3 shows that counter-intuitively the frequency of occurrence of the four different outcomes does not change dramatically relative to the case in which entry is open to all.

Latitude of rejection	Open entry		Selective entry		Scenario
	Occurrence	*Turnover*	*Occurrence*	*Turnover*	
Low	99.7%	46.9%	100.0%	48.7%	*Forking polarization*
	0.0%	-	0.0%	-	*Parallel polarization*
	0.0%	-	0.0%	-	*Winner-take-all*
	0.3%	0.0%	0.0%	-	*Consensus*
High	39.7%	37.9%	45.7%	43.9%	*Forking polarization*
	11.3%	0.0%	11.5%	0.0%	*Parallel polarization*
	3.1%	0.0%	2.8%	0.0%	*Winner-take-all*
	45.9%	0.0%	40.0%	0.0%	*Consensus*

Tableau 3 - Frequency of occurrence of the four possible outcomes and turnover rate

The data used to produce the table are averages taken over ten simulation runs for each combination of lat^r and lat^a. All other parameters are as above. Reading: Frequency of occurrence of the four possible outcomes at the end of a simulation run, and turnover rate (measured as number of replaced agents per 100 time-steps), for different ranges of the latitude of rejection lat^r (low: up to 1.1, high: 1.2 and higher) when exit is allowed and entry is open to agents with any orientation (left) or restricted to agents with positive orientation (right). For example the first line, first column indicates that with low lat^r, Forking Polarization occurs almost all the time and every 100 new messages, about 47 agents leave the forum and are replaced by newcomers with any orientation (pro-pathology or pro-recovery); the first line, second column indicates that these figures do not change substantially if the newcomers have all positive (pro-recovery) orientations.

The turnover rate sheds light on this (otherwise surprising) result. In some cases, a stationary state is reached without inducing any exit and therefore, without any entry; a policy of replacing leavers with a particular type of selected agents would thus never kick in. It is practically only when the final outcome is Forking Polarization, that turnover is high and therefore, the injection of agents with a specific orientation may have some effect. However, this effect consists only in creating an asymmetry between the two clusters of agents that emerge in forking cases—one grouped around one extreme, the other around its opposite. When entry is open, about 50% of agents are on either extreme of the admissible range; when it is selective, about 67% of agents cluster around the "positive" extreme, which we interpret as a strong pro-recovery orientation, while the remaining 33% form a group around the "negative" extreme, which we interpret as the more pro-pathology orientation. Turnover is higher because the presence of positively inclined newcomers increases the likelihood of departure of negatively inclined agents. The population changes through replacement by newcomers, but rarely extinguishes the negative pole.

In sum, in case of selective entry, greater presence of orientations of one type does not drive the whole forum to sharing its orientation; influence remains limited and dissenting agents remain, though fewer in number, without renouncing their convictions.

Effects of structural factors: Rate of participation

What happens when members of the forum are more active in discussions, and in particular, in replying to, or commenting on, messages posted by other members? Table 4 presents the effects of different levels of the rate of participation, operationalized as the percentage of forum members who post replies to a message by one of them. How do the results presented above vary if 10%, 40%, 70% or 100% of members post replies to each message initiating a thread? Table 4 shows the results for all possible combinations of $lat^r$and lat^a, comparing cases in which exit only is allowed, and cases in which entry is allowed too. The other parameters remain as before.

	Latitude of rejection	Rate of participation				Scenario
		10%	*40%*	*70%*	*100%*	
Exit only	*Low*	0.0%	0.0%	0.0%	0.1%	*Forking polarization*
		0.0%	0.0%	0.0%	0.0%	*Parallel polarization*
		100.0%	99.9%	99.4%	99.9%	*Winner-take-all*
		0.0%	0.1%	0.6%	0.0%	*Consensus*
	High	2.3%	2.3%	1.8%	1.3%	*Forking polarization*
		10.9%	12.2%	13.6%	13.6%	*Parallel polarization*
		44.5%	46.9%	50.2%	51.3%	*Winner-take-all*
		42.3%	38.6%	34.4%	33.8%	*Consensus*
Exit and Entry	*Low*	99.7%	99.9%	98.6%	98.6%	*Forking polarization*
		0.0%	0.0%	0.0%	0.0%	*Parallel polarization*
		0.0%	0.1%	0.1%	0.4%	*Winner-take-all*
		0.3%	0.0%	1.3%	0.9%	*Consensus*
	High	39.7%	42.6%	36.4%	36.7%	*Forking polarization*
		11.3%	12.2%	12.2%	12.5%	*Parallel polarization*
		3.1%	7.8%	11.7%	14.6%	*Winner-take-all*
		45.90%	37.4%	39.7%	36.2%	*Consensus*

Tableau 4 - Frequency of occurrence of the four possible outcomes, by participation rate

The data used to produce the table are averages taken over ten simulation runs for each combination of lat^r, lat^a, and the rate of participation, in cases in which Exit only, then Exit and (non-selective) Entry are allowed. All other parameters are as above. Reading: Frequency of occurrence of each possible final outcome, in percentage, for different ranges of the latitude of rejection lat^r(low: up to 1.1, high: 1.2 and higher), when participation is 10%, 40%, 70% and 100%. For example, the bottom row indicates that when both entry and exit are allowed, and $lat^r$is high, Consensus is observed about 46% of the time when only 1 out of 10 forum members replies to a message, and less often, i.e., only about 36% of the time, when all forum members reply.

Winner-Take-All always remains dominant for lower values of lat^r(at or below 1.1). Very rarely, other outcomes may be observed when participation is high (70% or above). For higher lat^r(at or above 1.2), Winner-Take-All and Consensus are about as frequently observed when 10% of forum members reply to messages; Winner-Take-All becomes increasingly frequently observed, and Consensus increasingly less frequently, with higher levels of participation rate. The frequency of occurrence of polarization (both Forking and Parallel) does not change with the rate of participation.
The remaining size of the population in stationary state, and the time needed to achieve it, illuminate the social mechanisms underpinning this result. When the outcomes are Consensus, Forking or Parallel Polarization, convergence time is always very short regardless of the participation rate; however, when the outcome is Winner-Takes-All, higher participation rates require longer time to convergence. With higher rates of participation, it takes longer to reach a stationary state, all the more so as the latitude of rejection is low (up to 1.1). Further, in Winner-Take-All scenarios with 10% participation, the size of the community in stationary state is usually significantly smaller than that of the original population (10% to 30% of it, depending on latitude of rejection); higher rates of participation correspond to a larger number of remaining agents (20% to 50% of the original population size, depending on latitude of rejection, with 100% participation).
It follows that higher rates of participation allow more influence to take place in the system; members take longer time to discuss, and many more eventually adhere to the orientations of the majority and refrain from leaving. The Winner-Take-All outcome is achieved in longer time, with a larger remaining population, because many accept to change their orientations rather than exiting the forum.

The dynamics of conflict and support

To better understand the social mechanisms through which the individual characteristics and structural factors explored above affect the final outcomes of the system, it is useful to look more closely at the progression of a single simulation run and in particular at the number of supportive and conflictive replies that each message elicits, under different parameter values and initial conditions. Replies play a very important role both in prompting agents to exit (when conflictive replies consistently and repeatedly outnumber supportive ones, as determined by the patience parameter) and in influencing their orientations (as agents revise their views in reaction to the collective, supportive and conflictive, replies of others). The effects of parameters on the amount and content of replies is thus essential to explain the reactions of agents, in terms both of influence on their orientations and of decisions to remain or leave the forum.
It appears that, with low levels of $lat^r$and no turnover, conflictive replies dominate and at the end, are twice as numerous as supportive ones. This is one reason why Forking Polarization ensues in these cases, with division of agents into two opposing camps. Instead, when there is exit but agents who leave are not replaced, supportive and conflictive replies are about equal in number over a simulation run, but are unequally distributed over time: the number of conflictive replies first outweighs supportive ones, until all misaligned agents exit the

community; at this stage all remaining agents have close enough orientations, there is hardly any scope for more conflictive replies, and supportive ones come to dominate until the Winner-Take-All stationary state is reached. The high number of conflictive replies at the beginning explains the high exit rate and the fact that the resulting Winner-Take-All configuration occurs with a very small remaining community. If leaving agents are replaced by newcomers, conflictive replies dominate again and Forking Polarization becomes the most likely outcome.

For higher levels of lat^r, the number of supportive replies increases with lat^a and in particular tends to be higher than the number of conflictive ones; supportive replies tend to overwhelmingly prevail, for almost all levels of lat^a above 1.2. As discussed above (Figure 5), in these cases more moderate orientations tend to emerge at the end, whether in the form of Consensus or of Parallel Polarization, under all three conditions of turnover.

In turn, higher rates of participation tend to exacerbate these patterns, because more agents reply to each single message posted in the forum. With both exit and (non-selective) entry, higher rates of participation increase the ratio of conflictive vs. supportive replies when the outcome is Forking Polarization (and, to a lesser extent, Winner-Take-All), and decrease it when the outcome is Consensus.

Technical comparison with agent-based opinion models

To fit the functioning of online forums, our model distinguishes itself from the literature that inspired it. We integrate the Jager-Amblard (2005) representation of influence into an original structure of interaction, in-between G. Deffuant et al. (2002) and R. Hegselmann and V. Krause (2002). Our focus is also different, less keen on identifying the effects of individual characteristics (latitude of acceptance and rejection) and more on structural factors (turnover and participation), which could be observed empirically and are amenable to policy action. Turnover in particular, has been little studied before, apart from a previous model built by two of the present authors (Rouchier, Tubaro and Emery 2014), and a study considering exit (Biswas, Sinha and Pen 2013) in the very distinct context of a voter problem. A model more similar to ours is built by M. Pineda, R. Toral and E. Hernandez-Garcia (2009, 2011, 2013) whose variation from the G. Deffuant et al. (2002) and R. Hegselmann and V. Krause (2002) approaches consists in adding randomly changing opinions in agents, and in allowing them to leave the group (not necessarily due to dissatisfaction as in our case). These features produce more populated moderate subgroups, and reduce the size of extreme groups.

Interestingly, we observe a smaller range of possible equilibrium orientations (one or two depending on parameters) than in all these other models, such as the one of W. Jager and F. Amblard (2005) which produces up to 6 orientation values when latitudes increase. The simulations of M. Pineda and co-authors also yield a larger number of stable outcomes at the end. These models all assume dyadic interactions or very global communication structures: it can thus be inferred that the specific properties of online forum interactions, in-between the two, that we have implemented here, drive greater convergence in orientations.

Discussion

The ways conflict and support affect the four possible equilibriums of the health and eating online forums at the center of our study are manifold, both at the level of single interactions (expressed by hostile/disagreeing or helpful/agreeing replies) and at a more macro level (emerging from the resulting turnover in each forum). They have significant impact on the results of our model and allow us to discuss our hypotheses.

H1 confirmed: Consensus prevails when supportive replies prevail among a population without turnover

Support plays an important role in determining when ambivalent and opposing orientations coexist within a forum. As noticed above, when the latitude of rejection allows sufficient tolerance towards conflicting orientations ($1.2 = lat^r = 2$), Forking Polarization gives way to other equilibriums. All four scenarios can come into existence, whether they correspond to lesser support (Forking Polarization and Winner-Take-All, where mean o takes more extreme values) or more (Parallel Polarization and Consensus, where o takes more moderate values on average).

Under which conditions do the latter two scenarios prevail? Specifically, how can we reach Consensus, a single common orientation characterized by more support and more moderate values of o? This is a test for our first hypothesis (H1). The findings summarized in Figure 5 provide some possible answers. When both the latitude of rejection lat^r and the latitude of acceptance lat^a are sufficiently high, orientations stop being ambivalent and converge towards a common moderate level. But what exactly brings about this equilibrium? The ratio of supportive/ conflictive replies shows that for higher levels of lat^r and lat^a and no exit allowed, virtually no conflictive reply is posted—only supportive ones. That moderates the general orientation of the community. On the contrary, for smaller values of lat^r, fewer messages are exchanged overall. This leaves the system in a situation of indecision and ambivalence, where no middle ground between antagonistic orientations can be found.

H2a confirmed: A single orientation prevails when dissenting agents are allowed to leave—not because of influence

The case of low lat^r corresponds to overwhelming dominance of one outcome: Forking Polarization or Winner-Take-All, both involving convergence to a non-moderate orientation value. The simulations allowing only exit show that this is due to the fact that those agents who do not share common orientations just leave the forum: the winner takes all not because it influences the other agents, but because it drives them away. We can conclude that social influence has a small effect in this case, and exit is the sole pertinent social dynamics as the agents reserve the right to leave the community if the overall orientation does not fit with their own.

This is quite manifest when we analyze the ratio supportive/conflictive replies over time. When exit is allowed and for smaller lat^r values, the Winner-Take-All equilibrium kicks in only once all the conflictive replies stop. In this case, even if the overall number of conflictive replies is close to that of supportive ones, the former are more abundant but occur over a shorter time, the latter are less abundant but are observed for longer—they go on even after all the conflictive replies have stopped. Conflictive and supportive replies have different temporalities. With this configuration of parameters, conflict happens at the beginning of the simulation; subsequently, disagreeing agents walk out on their fellow community contributors. This confirms H2a: one single orientation prevails because forum members who are in disagreement are progressively excluded. The end state is a scenario of radicalization, although no indication is provided as to whether the final orientation of the remaining agents would be a pro-pathology or a pro-recovery one.

H2b confirmed: When members are allowed to freely join or leave a population, more dissension is expressed and outcomes are more polarized

More generally, turnover is a focal element in our model, as the possibility of joining or leaving a community is part of the experience of Internet users and contributes to the social dynamics taking place online. Our agent-based simulations suggest that this has an effect on the appearance of shared orientations. We have proposed a comparison among three possible situations: no turnover allowed, only exit allowed, exit and entry allowed. Figure 5 highlights that in the latter case, with weaker acceptance of diverging orientations ($0 < lat^r \leq 1.1$), the system does not converge towards a situation of Winner-Take-All. On the contrary, it becomes ambivalently polarized on two extreme and Forking orientations. This confirms H2b. In this hypothesis, we supposed that agents exposed to conflict would simply decide to move to a place where their orientations would not be challenged. This is true, but they are replaced by other agents who can join the dissenting group and maintain it. The final situation is a polarized one.
In the specific context of our research, this has a positive consequence. As mentioned earlier, some more radical contributors to online communities about eating disorders are often accused of exposing others to harmful messages and proselytizing their pro-ana orientations. Our results show that influence is not effective, and that the risk of proselytism is extremely feeble.
This is confirmed by the fact that selective entry, allowing only agents with a specific orientation to join an existing online community, does not change the overall frequency of occurrence of each scenario in comparison to an open entry mechanism (Table 3). The global patterns remain the same, except the percentage of community members sharing that same orientation at the end of the simulations in the case of Forking Polarization—which is actually higher. This, again, confirms that influence is not running in this setting: the injection of pro-recovery agents does not succeed in changing other members' orientations.

H3 is only partially confirmed: If more agents actively participate in the interaction, social influence operates— but this only achieves weak polarization

The only context where influence seems to operate some minor changes in our findings is when we increase the ratio of participation in an online forum. At first

glance, the structure of our results in terms of frequency of occurrence of each scenario for higher and lower values of lat^r does not change dramatically. H3 is only (very partially) confirmed: for lower lat^r values, a very weak occurrence of Forking Polarization is detectable, while for higher lat^r values there is a slight increase in Parallel Polarization. But at a closer look, the situation is more multifaceted. Agents have to interact longer in order for the community to reach an equilibrium, and more participation entails a larger size of the forum in terms of remaining agents at the end of each simulation for every value of lat^r and for all possible scenarios.
Participation may also affect the extent to which social influence operates in the system. As far as more participation in our simulated community translates into more forum members interacting for a longer period of time, influence should be observed more frequently. Although the Winner-Take-All equilibrium normally prevails, indicating that community members often leave rather than convert to the pro-recovery orientation, high participation and selective entry together may bring about improvements.

* *
*

Conclusions and policy recommendations

The interest of our agent-based model lies in the comparison of possible empirically-informed scenarios to assess how different structural factors shape the social functioning of a human collective. Because they focus on healthcare, the specific computer-mediated social interactions we have taken into consideration are situated at the crossroads of intimate motives, group dynamics and macro-level concerns. It is therefore suitable to ask what recommendations this study can offer to policymakers and healthcare professionals.
It goes without saying that no social science approach can provide a precise recipe for success, and the limitations of the methods we have used—qualitative fieldwork and computer simulation—are well known. Further, as in each modeling effort, we have had to simplify and schematize what is in fact a very complex social process, and we have focused only on some particular aspects of the dynamics of orientation formation, to ensure tractability. Yet by combining qualitative and computational insights, we have been able to frame questions in a new way, and obtain novel results that could not obviously be expected ex ante.
The main lesson learned is that the sheer presence of contents challenging medical mediation in online interactions centered on eating disorders does not inevitably lead to a generalization of pro-pathology stances. Both the fieldwork and the simulation support this result, shedding new light on the perceived risk of eating disorders-related online interactions and inviting to reconsider the associated moral panic. A more comprehensive policy approach seems better adapted to these findings.
Firstly, our in silico analysis of eating disorder online communities puts in critical perspective some of the policies already attempted by governments and Internet service providers. Since 2001, there has been a succession of ineffective

bans, web filtering and censorship of eating disorder-related web contents, as well as proposals to enact restrictive legislation, but the number and popularity of anorexic and bulimic communities have not shrunk (Casilli, Pailler and Tubaro 2013). Our agent-based model provides suggestive reasons for the meager results of policy measures preventing entry to such communication platforms. Restricting access brings forth scenarios with less support: Forking Polarization if exit is not possible, and Winner-Take-All if it is. The former case being not coherent with the observed social phenomenon which establishes the empirical basis for our model (where exit from an online interaction is always a possibility), we can conclude that if entry is prevented by censorship or web filtering, communities tend to converge around more extreme health orientations. We have shown that this is no guarantee that pro-recovery stances become dominant. This would depend on the general level of acceptance $lat^a$and rejection lat^r, two parameters related to cultural traits over which policymakers have little or no direct control.

It is also worth reminding that, given the limited effectiveness of social influence in changing agents' orientation, Winner-Take-All scenarios emerge not because forum members are convinced to change their attitude towards medical mediation, but because they are forced out of the community by disagreeing agents and conflictive replies. Consequently, although online public health campaign designs that leverage influence mechanisms have been widely recommended, their effectiveness is uncertain in our specific case. Even network health interventions (Valente 2012) based on the selective entry of influencers (such as pro-recovery "ambassadors" or health professionals joining the online communities) may only succeed in crowding out, not in winning over, agents with diverging stances.

Nevertheless, an intervention consisting in selectively encouraging pro-recovery agents to enter the forum, coupled with a community management policy furthering participation, may expose forum members to health messages for longer and increase the odds of yielding pro-recovery Winner-Take-All equilibriums without excessively reducing the forum population.

Overall, this curbs alarmist claims about the risk of "pro-ana discourse going viral on the Internet." The sheer presence of web users bearing pro-ana messages will not automatically propel the generalization of pro-pathology orientations— pace policymakers. Notice, however, that this result also applies to the opposite orientation, as possible pro-recovery influence and proselytizing—even if systematically put in place—would also turn out to be less effective.

Possibly, the best guideline would be to surrender unrealistic policy goals in terms of absolute pro-recovery orientation, and to work on consensus building around more moderate orientations. This would mean accepting the free flow of orientations in forums, and relinquishing some of the prerogatives healthcare professionals have traditionally guarded as gatekeepers of medical information. Today's health professionals and online patients are both part of a complex ecosystem of "apomediaries" (Eysebach 2008) who "guide from a distance" (the Greek prefix apo- implying remoteness) other agents in selecting health services via recommendation, discussion, cooperation. But, unlike in conventional medical mediation, they do not achieve this by exerting a power to alter or select information and access to resources.

APPENDIX

Appendix 1: Model formalization

What follows is a more detailed description of the features of the model, which has been programmed and run with NetLogo 5.0.4 (Wilensky, 1999).

a) An *agenti* in the model is defined by four attributes: orientation $o_i(t)$, latitude of acceptance lat^a, latitude of rejection lat^r and patience p, where t designates a point in time. Only orientation is an individual value. All three other values are identical for all agents and do not change over the simulation. At initialization, $o_i(0)$ is randomly attributed to the agent following a uniform distribution on [-1; 1]. lat^a and lat^r are defined on [0; 2[for lat^a and on $]lat^a; 2]$ for lat^r; p is an integer, greater than 0 and equal for all agents.

b) A time step is defined as:

b1) a complete topic thread among agents

A *topic thread* is such that one single agent i (the original sender) is chosen randomly from the population and sends an original message ($mess\ (t)$) that contains information $o_i(t)$ about its orientation ($mess\ (t) = o_i(t)$).

The *perception* of another agent j is defined by $o_j(t)$, lat^a and lat^r: there is an interval of agreement in which the agent agrees with the message ($[o_j(t)- lat^a;\ o_j(t) + lat^a]$) and an interval of disagreement ($[\ -1\ ;\ o_j(t) - lat^r] \cup [\ o_j(t)+ lat^r; 1]$).

Among all the agents j who perceive the message, a subset will react by sending the reply rep_j, containing their own orientation.

This subset is a proportion defined by the *rate of participation* of the model:

- If mess (t) ∈ [$o_j(t)$ - lat^a; $o_j(t)$+ lat^a] then $rep_j(t) = o_j(t)$ and $rep_j(t)$ is "supportive";
- If mess (t) ∈ [-1 ; $o_j(t)$ - lat^r] ∪ [$o_j(t)$+ lat^r; 1] then $rep_j(t) = o_j(t)$ and $rep_j(t)$ is "conflictive".

The set of supportive replies is $Set^s(rep^s(t))$and the set of conflictive replies is $Set^c(rep^c(t))$. $Avg^s(rep^s(t))$and $Avg^c(rep^c(t))$ are the average values of orientations of (resp.) all replies in $Set^s(rep^s(t))$ and $Set^c(rep^c(t))$:

- $Avg^s(rep^s(t))$ = (∑ (in $Set^s(rep^s(t))$ $rep_j(t)$)) / (size $Set^s(rep^s(t))$)
- $Avg^c(rep^c(t))$ = (∑ (in $Set^c(rep^c(t))$ $rep_j(t)$)) / (size $Set^c(rep^c(t))$)

b2) a change of orientation for all agents, linked to the interaction that took place

A *change of orientation* takes place after all potential messages (original message and replies) in a given topic thread are exchanged. Any agent, whether it partook in the thread or not, who was able to perceive $Avg^s(rep^s(t))$ and/or $Avg^c(rep^c(t))$, modifies its orientation accordingly in two steps, first by being influenced by supportive messages, then by conflictive ones (we thus introduce a local variable, *temp*). An agent who only perceives $Avg^s(rep^s(t))$ (resp. $Avg^c(rep^c(t))$) is influenced once; an agent who perceives none of these averages does not modify its orientation at that step:

- If $Avg^s(rep^s(t))$ ∈ [$o_j(t)$- lat^a; $o_j(t)$+ lat^a] then $temp = o_j(t) + \mu(Avg^s(rep^s(t)) - o_j(t))$;
- If $Avg^s(rep^s(t))$ ∈ [-1; $o_j(t)$ - lat^r] ∪ [$o_j(t)$+ lat^r; 1] then $temp = o_j(t) - \mu(Avg^s(rep^s(t)) - o_j(t))$;
- If $Avg^c(rep^c(t))$ ∈ [$o_j(t)$- lat^a; $o_j(t)$ + lat^a] then $o_j(t+1) = temp - \mu(Avg^c(rep^c(t)) - temp)$;
- If $Avg^c(rep^c(t))$ ∈ [-1; $o_j(t)$ - lat^r] ∪ [$o_j(t)$+ lat^r; 1] then $o_j(t+1) = temp + \mu(Avg^c(rep^c(t)) - temp)$.

Table A1 below provides a numerical example of this process, based on the values of the orientations of five agents as illustrated in figure 2 (above), and assuming that lat^a = 0.3, lat^r = 0.6. Agent *i* initiates the thread by posting a message equal to its orientation, 0.12; this value falls within the latitude of acceptance of agents *k* and *l* who express their support, while it is in the latitude of rejection of agent *h* who replies unsupportively, and it is in the neutral zone of agent *j* who does not react. The average value of all supportive replies $Avg^s(rep^s(t))$ is the mean of the orientations of *k* and *l*, equal to 0.025, while the average of conflictive replies is just the one of agent *h* and is equal to its orientation, - 0.8. This information enables to compute the change of orientations for all agents, first by calculating *temp* and then by deriving the new value of their orientation, according to the above formulas. In this case, the agent who initiated the trend had a relatively moderate orientation, and so did its supporters, but there was one very strong reaction; as a result, all orientations are driven further away from positive (pro-recovery) and down towards more negative (pro-pathology) stances.

Agent	$o(t)$	$o(t) - lat^a$	$o(t) + lat^a$	$o(t) - lat^r$	$o(t) + lat^r$	Type of reply	*temp*	$o(t+1)$
i	0.12	-0.18	0.42	-0.48	0.72	(initiates)	0.1105	0.0195
k	0.15	-0.15	0.45	-0.45	0.75	supportive	0.1375	0.0438
l	-0.1	-0.4	0.2	-0.7	0.5	supportive	-0.0875	-0.1588
h	-0.8	-1	-0.5	-1	-0.2	conflictive	-0.8825	-0.8907
j	0.5	0.2	0.8	-0.1	1.1	neutral	---	0.37

Table A1 – A numerical example of the interaction and influence process in a time step.

b3) potential exits and entries, in the cases when the system allows them.

After receiving all replies, the original sender evaluates them and decides to remain in the community or to quit. The agent counts the number of steps *NR* in a row when:

$$\text{size } Set^c(rep^c(t)) - \text{size } Set^s(rep^s(t)) > 0$$

and if $NR \geq p$ then it quits the community.

Entry of new agents is such that whenever one agent exits, it is replaced by another agent with random orientation and the same lat^a, lat^r and p as the others. Effective from the following time step, the newcomer can be the sender of an original message.

APPENDIX 2. – List of parameters and notations

Definition	Notation
Agents	i, j, l, k, h
Time	t
Orientation (of agent i at time t)	$o_i(t)$
Latitude of acceptance	lat^a
Latitude of rejection	lat^r
Patience	p
Message (at time t)	$mess\ (t)$
Reply (of agent j at time t)	$rep_j(t)$
temp	*temp*
Set of supportive replies (at time t)	$Set^s(rep^s(t)$
Set of conflictive replies (at time t)	$Set^c(rep^c(t)$
Average supportive replies (at time t)	$Avg^s(rep^s(t))$
Average conflictive replies (at time t)	$Avg^c(rep^c(t))$
Rate of influence	μ
Number of consecutive time steps in which there are more conflictive than supportive replies	NR
Average positive orientations (in stationary state)	$avg(o_{pos})$
Average negative orientations (in stationary state)	$avg(o_{neg})$
Rate of participation	---

Table A2. – List of parameters and notations for the model

REFERENCES

Agar M. H. (2003) My kingdom for a function: Modeling misadventures of the innumerate. *Journal of Artificial Societies and Social Simulation* 6, 3, URL http://jasss.soc.surrey.ac.uk/6/3/8.html

Akrich M. & Méadel C., 2007. De l'interaction à l'engagement : les collectifs électroniques, nouveaux militants dans le champ de la santé. *Hermès*, 47 (1): 145-153.

Amblard F., Bommel P., Rouchier J., 2007. Assessment and validation of multi-agent models. In *Multi-Agent Modelling and Simulation in the Social and Human Sciences*, Phan D. & Amblard F. (eds), Bardwell Press, pp. 93-114.

Axelrod R., 1997. The dissemination of culture. A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41: 203–226.

Biswas S., Sinha S. and Sen P., 2013, Opinion dynamics model with weighted influence: Exit probability and dynamics, arXiv:1306.6813.

Casilli A.A., 2013. Le Web des troubles alimentaires. Un nouvel art de jeûner? *Communications*, 92 (1): 111-123.

Casilli A.A., Pailler F., Tubaro P., 2013. Online networks of eating-disorder websites: why censoring pro-ana might be a bad idea. *Perspectives in Public Health*, 133 (2): 94-95.

Chattoe, E. (2002) 'Building empirically plausible multi-agent systems: A case study of innovation diffusion', in Dautenhahn K., Bond A.H., Cañamero D. and Edmonds B. (eds.) *Socially Intelligent Agents: Creating Relationships with Computers and Robots, Multiagent Systems, Artificial Societies and Simulated Organisations,* Volume 3 (Dordrecht: Kluwer), pp. 109-116.

Deffuant G., Amblard F., Weisbuch G., Faure T., 2002. How can extremism prevail? A study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5, 4, URL: http://jasss.soc.surrey.ac.uk/5/4/1.html

Durkheim É., 1893. *De la division du travail social.* Paris: PUF.

Epstein J. M. (2002) Modeling civil violence: an agent-based computational approach. *Proceedings of the National Academy of Sciences of the United States of America*, n. 99 Suppl 3: 7243-7250.

Eysenbach G., 2008. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *Journal of Medical Internet Research*, 10(3): http://www.jmir.org/2008/3/e22/

Eysenbach G., 2004. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions, *British Medical Journal* 328 (7449): 1166.

Galam S., Moscovici S., 1991. Towards a theory of collective phenomena: consensus and attitude changes in groups. *European Journal of Social Psychology*, 21: 49-74.

Gilbert N., Troitzsch K, 1999. *Simulation For The Social Scientist*. McGraw-Hill International.

Grimm V., et al. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 198 (1-2): 115-126.

Hargittai E., Walejko G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society* 11(2): 239 256.

Hegselmann R., Krause U., 2002. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5, 3, URL: http://jasss.soc.surrey.ac.uk/5/3/2.htm

Jager W., Amblard F., 2005. Uniformity, bipolarisation and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational and Mathematical Organization Theory*, 10: 295-303.

Kozma B., Barrat A., 2008. Consensus formation on coevolving networks: Groups formation and structure. *Journal of Physics A: Mathematical and Theoretical*, 41: 224020.

Manzo G., 2007a. Variables, mechanisms, and simulations: can the three methods be synthesized? A critical analysis of the literature. *Revue Française de Sociologie*, 48 (Supplement *An Annual English Selection*): 35-71.

Manzo G., 2007b. Progrès et 'urgence' de la modélisation en sociologie. Du concept de modèle générateur et de sa mise en œuvre. *L'Année Sociologique*, 57(1): 13-61.

Nielsen J., 2006. *Participation Inequality: Encouraging More Users to Contribute*. URL: http://www.egov.vic.gov.au/website-practice/usability/participation-inequality-encouraging-more-users-to-contribute.html [last accessed Nov. 4, 2013].

Ochoa X., Duval E., 2008. Relevance ranking metrics for learning objects. I*EEE Transactions on Learning Technologies*, 1(1): 34-48.

Pineda M., Toral R., Hernandez-Garcia E., 2013. The noisy Hegselmann-Krause model for opinion dynamics, *European Physical Journal B*, 86: 490.

Pineda M., Toral R., Hernandez-Garcia E., 2011. Diffusing opinions in bounded confidence processes, *European Journal of Physics D*, 62, (109-117).

Pineda M., Toral R., Hernandez-Garcia E., 2009, Noisy continuous-opinion dynamics, *Journal of Statistical mechanics*, P08001, (1-18).

Ren Y., Kraut R., 2014. Agent-based modeling to inform online community design: Impact of topical breadth, message volume, and discussion moderation on member commitment and contribution. *Human Computer Interaction,* 29(4): 351-389.

Rouchier J., Tanimura E., 2012. When overconfident agents slow down collective learning. *Simulation*, 88: 33-49.

Rouchier J., Tubaro P., 2011. Can opinion be stable in an open network with hierarchy? An agent-based model of the Commercial Court of Paris. *Procedia-Social and Behavioral Sciences*, 10: 123-131.

Rouchier, J. Tubaro, P. and C. Emery (online first 2013), Opinion transmission in organizations: An agent-based modeling approach. *Computational and Mathematical Organization Theory*, DOI10.1007/s10588-013-9161-2

Schweitzer F., Garcia D., 2007. An agent-based model of collective emotions in online communities, *CCSS Working paper series*, CCSS – 10 – 007, http://arxiv.org/abs/1006.5305

Simmel G., 1908. *Soziologische Untersuchungen über die Formen der Vergesellschaftung,* Berlin: Duncker & Humblot.

Tubaro P., Casilli A.A., 2010. 'An ethnographic seduction' : how qualitative research and agent-based models can benefit each other. *Bulletin de Méthodologie Sociologique*, 106(1): 59-74.

Valente T.W., Fosados R., 2006. Diffusion of innovations and network segmentation: The part played by people in promoting health. *Sexually Transmitted Diseases*, 33(7): S23–S31.

Valente T.W., 2012. Network interventions. *Science*, 337 (6090): 49-53.

Van Rooy D., 2012. A connectionist ABM of social categorization processes.*Advances in Complex Systems*, 15(06), 1250077-1-1250077-21.

Wilensky U., 1999. *NetLogo*. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Yang L., Gilbert N., 2008. Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in Complex Systems* 11: 175–185.

Yeshua-Katz D., Martins N., 2013. Communicating stigma: The pro-ana paradox. *Health Communication* 285(5): 499-508.

Zhou T., Medo M., Cimini G., Zhang Z., Zhang Y., 2011. Emergence of scale-free leadership structure in social recommender systems. *PLoS ONE* 6 (7): e20648.

---

[i] The fieldwork has been conducted as part of the study The “Ana-Mia” Sociability: An Online/ Offline Social Networks Approach to Eating Disorders (ANAMIA, ANR-09-ALIA-001).

[ii] For a presentation of the model based on data visualization, see the companion web page: http://anamia.fr/en/research-areas/modelling-proanaforum/.

[iii] A similar iterative mechanism, allowing comparison of observed and simulated regularities but using quantitative data, is described in Gianluca Manzo (2007b).