

# An agent-based model of collective emotions in online communities

F. Schweitzer<sup>a</sup> and D. Garcia

Chair of Systems Design, ETH Zurich, Kreuzplatz 5, 8032 Zurich, Switzerland

Received 28 June 2010 / Received in final form 5 September 2010

Published online 6 October 2010 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2010

**Abstract.** We develop an agent-based framework to model the emergence of collective emotions, which is applied to online communities. Agent's individual emotions are described by their valence and arousal. Using the concept of Brownian agents, these variables change according to a stochastic dynamics, which also considers the feedback from online communication. Agents generate emotional information, which is stored and distributed in a field modeling the online medium. This field affects the emotional states of agents in a non-linear manner. We derive conditions for the emergence of collective emotions, observable in a bimodal valence distribution. Dependent on a saturated or a superlinear feedback between the information field and the agent's arousal, we further identify scenarios where collective emotions only appear once or in a repeated manner. The analytical results are illustrated by agent-based computer simulations. Our framework provides testable hypotheses about the emergence of collective emotions, which can be verified by data from online communities.

## 1 Introduction

How do collective phenomena arise from the interaction of many distributed system elements? This question is certainly at the heart of statistical physics. Over the last 150 years it has provided a large set of methodologies applied to physical systems, to infer from the properties of the elements on the micro level on the systems dynamics on the macro level. A very similar question is also asked in different other scientific disciplines. For example, in medicine one wishes to understand the reaction of the immune system based on the communication and coordinated action of e.g. B or T cells. In economics one is interested in the emergence of systemic risk [1] in a financial system based on the fault of firms or banks clearing their debts to other firms or banks.

To answer such questions, we need an appropriate description of the system elements, which are called agents in the following, and their interactions – but we also need an appropriate framework to predict from these ingredients the possible collective dynamics on the systems level. Without such a framework, we are only left with extensive computer simulations of multi-agent systems, in which, for given assumptions of the interactions, we have to probe the entire parameter space, to find out the conditions for certain collective phenomena.

In this paper, we want to develop such a framework to describe collective emotions in online communities. There

is no commonly accepted definition of collective emotions yet. According to [2], collective emotions are shared by large numbers of individuals, in contrast to group-based emotions that are felt by individuals as a result of their membership in a certain group or society. The former concept suggests that group members may share the same emotions for a number of different reasons, whereas the latter refers to emotions that individuals experience as a result of identifying with their fellow group members.

Hence, in our paper collective emotions are shared by a larger number of individuals as a result of both external events and nonlinear coupling between individuals. Similar to other collective states, also collective emotions can display new, emergent properties which cannot be traced back to individual contributions. Remarkably, the life time of a collective emotion is usually much larger than the one of an individual emotion. On the other hand, individual emotions show a different dynamics in the presence of collective emotions, simply because of the nonlinear feedback of the emergent collective emotion on the individual one.

In this general manner, collective emotions are not restricted to online communities. Instead they can emerge in any social context. The aim of our research, however, is to understand the dynamics in *online* communities. The cyberspace does not have emotions, but individuals that interact online can share emotions. Here certain conditions of individual interactions apply that are not present for offline communities. Online communities react on other time scales (not necessarily faster, but often with a time shift), they act on different stimuli (there are hardly seen offline

<sup>a</sup> e-mail: [fschweitzer@ethz.ch](mailto:fschweitzer@ethz.ch)

assemblies to share emotions about a Youtube video), they have different thresholds to express their emotion, and they do it in a very different manner, namely by writing in. For all these, we may gather data from online portals, whereas it becomes very difficult to measure those in offline communities. We remark that the internet is indeed shaping the phenomena in mind, it is not a mere interface for monitoring ‘real’ social interaction. While we agree that there are certain commonalities (mostly based on social herding and amplification), there are also substantial differences in communication. For the very same reason, we argue that real phenomena like mass hysteria can indeed be seen as instances of collective emotions, however, not all of the modeling implications used in this paper may apply.

Examples of collective emotions in online communities can be observed en masse on the Internet. One particular example was the large amount of emotional discussions which followed the death of Michael Jackson, other examples are the “memes” and heated discussions of anonymous fora like [4chan.org](http://4chan.org). They follow a very similar scheme: users which have subscribed to social network sites or to blogs or discussion fora, become enraged or excited about a particular event (like the performance of a beloved soccer team in a world competition) or a personal (good or bad) experience. Importantly, these individual feelings are then shared with other users, i.e. they are *communicated* by means of online media, most likely by writing a personal statement. Obviously, users do not transmit an emotion, instead they communicate a piece of information, which may trigger an emotional reaction in participants reading this. Dependent on such an impact, other users may decide to involve themselves in such an emotional communication, e.g. by sharing the feeling or opposing to it. Under certain circumstances, we may observe mutual communication in a small group of users, but there are also scenarios where many users express their feeling once, in a sequence, or where many users repeatedly fire the discussions by emotional statements. These discussions show the existence of emergent collective states in which the users share their emotions, rather than an aggregate of the emotions of the community. The discussions do not necessarily have to be centered around just one feeling. In many cases, we see the emergence of two collective emotions, a ‘positive’ and a ‘negative’ one, which may coexist or ‘fight’ each other. These collective states usually only have a finite life time, i.e. they disappear, but could come back when triggered by a new event or post.

How do we want to model such collective phenomena? In an agent-based model, we first need to describe the emotional states of individual agents, which should be based on insights obtained in psychology.

First of all, emotions are very different from e.g. *opinions* in that they are rather short-lived subjective states that decay to the neutral state very fast (see also the survey study of [3]). While both emotions and opinions can be influenced by herding effects, opinion dynamics is often linked to utilities and preferences, while emotions do not

need to follow any particular optimization space. Further, opinions are often becoming externalized instances on the collective level (the “public opinion” exist outside the individuals), whereas collective emotions are perceived in a rather implicit way. We recall that it takes sophisticated algorithms to extract them from blog entries, etc.

Secondly, emotions are characterized by different dimensions. An established theoretical perspective also used in this paper is the circumplex model [4] which is based on the two dimensions of *valence*, indicating whether the pleasure related to an emotion is positive or negative, and *arousal*, indicating the personal activity induced by that emotion. However, in the psychology literature various ways of representing emotions can be found (see [5,6] for a review of different dimensional representations). For the sake of completeness we mention that for computational models the appraisal theory [7] provides another promising theoretical perspective: it is based on internal representations of person-environment relations, which can be modeled by so-called BDI (belief-desire-intention) agents. While there is a large body of literature on computational appraisal models [8–10] the focus is more on the correct internal representation of emotions and their cognitive consequences, not on the explanation of collective phenomena such as described above.

Given that we are able to characterize agent’s emotions, how are we able to *detect* them? Internal emotional states can be inferred from physiological signals [11], in particular the internal dynamics and responses of humans in emotional contexts can be measured under the dimensional representation of [4]. In online communities, however, we cannot measure the physiological response of users directly. Instead, we are left with the problem to infer user’s emotions from the written text pieces they provide in the online media. Human annotation of internet data was used in [12], but this largely restricts the amount of data to be processed. Again, in computational sciences, there are established ways of *sentiment mining*, i.e. algorithms to extract the *emotional* content of a written text and to classify this according to various dimensions. Different sentiment classification techniques can be combined to improve results [13] and can be applied to study emotions in the internet [14]. Because in this paper we do not provide a direct comparison of our model results with empirical data, we skip the detailed discussion of those techniques, keeping in mind that we are indeed able to obtain e.g. the valences of different users participating in a blog, over time.

Another challenge results from the fact that we need to model the communication between users in online communities. It is not the emotion per se of an user what matters, but its expression in a blog entry, a post etc. This is submitted at a particular time and distributed to the whole online community, where it is perceived by other users with a very different time delay. While modeling a personalized communication would need to know the underlying social network, in most online communities a particular post is available to everyone (who has subscribed) immediately. This justifies the assumption of a mean-field

coupling between users, i.e. a medium is updated instantaneously and provides the same information to everyone. Nevertheless, we still have to consider that for example older posts have less impact on users than more recent ones and that positive and negative posts may be submitted with different frequencies. In this paper, we cope with these requirements by introducing an emotional information field generated by the users, which stores and distributes this information accordingly. This idea was already successfully applied in other communication models describing biological or social systems [15,16].

Eventually, we need to model the impact of the emotional information on online users. Unfortunately, the psychological literature does not provide much insight into this problem. Therefore, we are left with providing hypotheses about the feedback between the emotional information and the individual dimensions of arousal and valence. By proposing a very general non-linear feedback, for which different special cases are explored, we are able to derive conditions under which the emergence of collective emotions can be expected. Different scenarios can be obtained: either the repeated occurrence of collective emotions, or the one-time collective emotion. These results can be seen as testable hypothesis, which may be verified either by psychological experiments or by data from online communities.

In conclusion, in this paper we wish to derive a quite general modeling approach, to explore the conditions under which collective emotions may emerge from interacting emotional agents. Understanding the emergence of collective emotions certainly has an impact beyond online communities. They play for example a crucial role in resolving conflicts in societies [2]. Collective emotions are also important for the efficiency of working groups [17]. However, online communities provide a much better starting point for understanding collective emotions. First of all, there is a large amount of data available from these communities. Consequently, large scale emotions have been already studied for songs, blogs or political comments [18]. In addition to this, some peculiarities of information exchange between users on the internet are suspected to have an emotional origin. For example, the network of posts and comments in various blog sites [19] has a strong community structure that could be created by emotional discussions. The effect of emotions in creating and reshaping social contacts was also modeled in artificial social networks [20].

Secondly, collective emotions are fostered by internet communication because of (a) the fast information distribution, and (b) the anonymity of users in the internet, which often seduces people to drift away from established norms and show a salient private personality. In fact, empirical studies [21] which compared the attitude change in virtual and face to face interactions, have demonstrated that human behavior and social norms are affected by internet interaction. Thirdly, the internet is seen as an important factor in defining present and future societies [22]. Collective emotions, such as hate, play an important role in the creation of collective identities. [23] provides a sys-

tematic study of collective identity in internet-based hate groups.

## 2 An agent-based model of emotions

### 2.1 The concept of Brownian agents

Our modeling approach is based on the concept of Brownian agents [16]. It allows to formalize the agent dynamics and to derive the resulting collective dynamics in close analogy to methods established in statistical physics. A Brownian agent is described by a set of state variables  $u_i^{(k)}$ , where the index  $i = 1, \dots, N$  refers to the individual agent  $i$ , while  $k$  indicates the different variables. These could be either *external* variables that can be observed from the outside, or *internal degrees of freedom* that can only be indirectly concluded from observable actions.

Noteworthy, the different (external or internal) state variables can change in the course of time, either due to influences of the environment, or due to an internal dynamics. Thus, in a most general way, we may express the dynamics of the different state variables as follows:

$$\frac{du_i^{(k)}}{dt} = f_i^{(k)} + \mathcal{F}_i^{\text{stoch}}. \quad (1)$$

This formulation reflects the *principle of causality*: any *effect* such as the temporal change of a variable  $u$  has some *causes* that are listed on the right-hand side. For the concept of Brownian agents, it is assumed that these causes may be described as a *superposition of deterministic and stochastic influences*, imposed on agent  $i$ . This distinction is based on Langevin's idea for the description of Brownian motion, which coined the concept. Hence, we sum up influences which may exist on a microscopic level, but are not observable on the time and length scale of the Brownian agent, in a stochastic term  $\mathcal{F}_i^{\text{stoch}}$ , while all those influences that can be directly specified on these time and length scales are summed up in a *deterministic* term  $f_i^{(k)}$ . This implies that the “stochastic” part does *not* impose any *directed* influence on the dynamics (which would have counted as deterministic), but on the other hand, it does not necessarily mean a white-noise type of stochasticity. Instead, other types such as colored noise, or multiplicative noise are feasible. Noteworthy, the strength of the stochastic influences may also vary for different agents and may thus depend on local parameters or internal degrees of freedom, as was already used in different applications [16]. The *deterministic* part  $f_i^{(k)}$  contains all specified influences that cause changes of the state variable  $u_i^{(k)}$ . This could be nonlinear interactions with other agents  $j \in N$  – thus  $f_i^{(k)}$  can be in principle a function of all state variables describing any agent (including agent  $i$ ). But  $f_i^{(k)}$  can also describe the response of an agent to available information, as it will be the case for cyberemotions. It should further depend on external

conditions – such as forces resulting from external influences (most notably information from mass media). Eventually,  $f_i^{(k)}$  may reflect an (external or internal) *eigendynamics* – in the considered case a relaxation of the excited emotional state of an agent (caused by saturation or exhaustion). In order to set up a multiagent system (MAS) we need to specify the relevant state variables  $u_i^{(k)}$  and the dynamics of their change, i.e.  $f_i^{(k)}$ , which means also to specify the interaction between the agents. We emphasize that the dynamics of the MAS is specified on the level of the individual agent, not on a macroscopic level, so the collective dynamics shall emerge from the interactions of many agents.

## 2.2 Emotional states

To quantify the emotional dynamics of an agent, we consider the following continuous variables, *valence*,  $v_i(t)$ , and *arousal*,  $a_i(t)$ . Both define a two-dimensional plane for the classification of emotions. Valence ( $x$ -axis) measures whether an emotion is positive or negative, arousal ( $y$ -axis) measures the degree of personal activity induced by that emotion. Hence, an emotional state is defined by  $e_i(t) = \{v_i(t), a_i(t)\}$ . For example, ‘astonished’ is an emotional state with both positive valence and arousal, ‘satisfied’ has a positive valence, but a negative arousal, ‘depressed’ has both a negative valence and arousal, and ‘annoyed’ has a negative valence and a positive arousal. We note that both valence and arousal describe *internal* variables, i.e. a dynamics inside the agent, which may be only indirectly observable, for example through physiological measurements.

Without any internal or external excitation, there should be no positive or negative emotion, so we assume that in the course of time both valence and arousal relax into an equilibrium state,  $e_i(t) \rightarrow 0$ , which implies  $v_i(t) \rightarrow 0$ ,  $a_i(t) \rightarrow 0$ . Hence, in accordance with equation (1) we specify the dynamics of the Brownian agent as follows:

$$\dot{v}_i = -\gamma_{vi} v_i(t) + \mathcal{F}_v + A_{vi} \xi_v(t) \quad (2)$$

$$\dot{a}_i = -\gamma_{ai} a_i(t) + \mathcal{F}_a + A_{ai} \xi_a(t). \quad (3)$$

The first term in each equation describes the relaxation into an equilibrium state as an exponential decay of both valence and arousal, if no excitation is given.  $\gamma_{vi}$ ,  $\gamma_{ai}$  define the time scales for this relaxation, which are different for valence and arousal and further may vary across individual agents. The second and third term in the equations above describe influences which may induce an emotional state. These can be stochastic influences, expressed by the third term, where  $\xi_v(t)$ ,  $\xi_a(t)$  are random numbers drawn from a given distribution of stochastic shocks, with the mean of zero  $\langle \xi(t) \rangle = 0$  and no temporal correlations between subsequent events  $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$ .  $A_{vi}$ ,  $A_{ai}$  denote the strength of these stochastic influences which may again vary across agents. The two functions  $\mathcal{F}_v$ ,  $\mathcal{F}_a$  describe deterministic influences which cause the emotional

state. They very much depend on the specific assumptions applicable to collective cyberemotions, in particular the agents’ interaction, access to information, response to the media, but can depend also on internal variables such as *empathy*, i.e. the ability to share the feelings of other agents, or responsiveness to available information. Most of all, these functions should also reflect a dependence on the emotional state itself, i.e. agents already in a specific mood may be more affected by particular emotions of others. Before we specify these functions in detail, we need to extend the agent description.

## 2.3 Emotional actions

The dynamics of equations (2), (3) already define a stationary state  $e_i(t) \rightarrow 0$ , given that the deterministic and stochastic influences become negligible. On the other hand, there should be an excited emotional state of the agent if these influences are large, e.g. if information with a large emotional content becomes available to the agent. Per se, this state is not observable from the outside unless the agent takes any action that *communicates* that emotional state, for example by posting in a blog, etc. Consequently, we assume that the agent expresses its *valence*, i.e. the good or bad feeling, *if its arousal*, i.e. the action induced by the emotion, exceeds a certain individual threshold,  $\tau_i$ :

$$s_i(t + \Delta t) = \text{sign}(v_i(t)) \Theta[a_i(t) - \tau_i]. \quad (4)$$

Here  $\Theta[x]$  is the Heavyside function which is one only if  $x \geq 0$  and zero otherwise. If  $\Theta[x] = 1$ , we make the simplifying assumption that the agent does not communicate all details about his feelings (i.e. the value of  $v_i$ ) because perfect emotional information cannot be communicated. Instead, the agent communicates only if it is a good or bad feeling, i.e. the sign of  $v_i$ ,  $-1$  or  $+1$ , which is defined as  $r_i(t) = \text{sign}(v_i(t))$  in the following (note: the model specified here is not really changed if indeed the  $v_i$  is communicated, but the analytical investigations become more involved). This coarse-grained description of the valence only enters the communication process, while valences are still distributed across agents.

Equation (4) further reflects the assumption that the agent does not immediately express its feelings if the arousal hits the threshold at time  $t$ , but probably with a certain delay  $\Delta t$ , which may be caused by the fact that the agent has no immediate access to some communication media (computers in the case of cyberemotions) or other things to do. More important feelings should be communicated with a shorter delay. It should vary as well across agents. In accordance with investigations of waiting time distributions in performing human activities (e.g. answering emails), we may assume that  $\Delta t$  can be random drawn from a power-law distribution  $P(\Delta t) \propto \Delta t^{-\alpha}$ , where  $\alpha$  should be empirically determined. Note that the dynamics of the external state variable  $s_i(t)$  differ from the form given in equation (1) in that the stochastic influences are



not additive, but implicitly present because of the stochastic dynamics for  $v_i(t)$  (determining the *sign* of the expression),  $a_i(t)$  (determining the *time* of the expression) and  $\Delta t$  (determining the *delay* of the expression).

Based in equation (4), we can define the number of emotional expressions at a given time  $t$  as

$$N_s(t) = \sum_i \Theta[a_i(t) - \tau_i]. \quad (5)$$

Assuming continuous time, the *average* number of expressions per time interval then results from

$$n_s = \frac{1}{t_{\text{end}}} \int_0^{t_{\text{end}}} N_s(t) dt. \quad (6)$$

We may calculate this quantity from analytical approximations in Section 3.3 and from computer simulations in Section 4.

### 2.4 Communicating emotions

By now we have described the (internal) emotional dynamics of an agent that leads to a certain (externally visible) expression of an emotional state. In order to describe cyberemotions as collective emotions, we now need to specify how this emotional expression is communicated to other agents. In accordance with previous investigations [15] we assume that every positive or negative expression is stored in a *communication field*  $h_+(t)$  or  $h_-(t)$  dependent on its value.  $h_{\pm}(t)$  represent the communication media available for the storage and distribution of emotional statements, for example blogs, forums, etc. and simply measure the ‘amount’ of positive or negative feelings available at a given time. For the dynamics of the field, we propose the following equation:

$$\dot{h}_{\pm} = -\gamma_{\pm} h_{\pm}(t) + s N_{\pm}(t) + I_{\pm}(t). \quad (7)$$

Each agent contribution  $s_i(t)$  increases the respective field  $h_+$  or  $h_-$  by a fixed amount  $s$  at the time of expression, which represents the impact of the information created by the agent in the information field, as a time scale parameter.  $N_{\pm}(t)$  is the total number of agents contributing positive or negative statements at a given time  $t$ , i.e. all the agents with  $s_i(t) = 1$  and with  $s_i(t) = -1$  respectively.

The relevance of contributions fade out over time as e.g. agents become less affected by old blog entries. This is covered by an exponential decay of the available information with the time scales  $\gamma_{\pm}$ . Eventually, in addition to the agent contributions, positive or negative emotional content from the news may add to the communication field, which is covered by an agent-independent term  $I_{\pm}(t)$ , which can be modeled for example by a stochastic input.

The main feedback loops of this framework are sketched in Figure 1, where we can distinguish between two layers: an internal layer describing the agent (shown horizontally) and an external layer describing the communication process (shown vertically). In the internal layer,

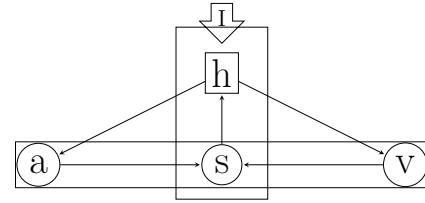


Fig. 1. Causation among the components of the model.

the arousal  $a$  and the valence  $v$  of an agent determine its emotional expression  $s$ , which reaches the external layer by contributing to the communication field  $h$ . The latter one has its independent dynamics and can, in addition to contributions from other agents, also consider input from external sources,  $I$ . The causality is closed by considering that both valence and arousal of an agent are affected by the communication field.

In order to complete the model, we need to specify how the available information affects the emotional states of the individual agents, which is covered in the functions  $\mathcal{F}_v$  and  $\mathcal{F}_a$ .

### 2.5 Emotional feedback

Because we are interested in the outbreak of *collective* emotion, we do not assume the latter as the simple superposition (or addition) of individual emotional states. On the contrary, we assume that an emotional state of one agent, if it is expressed and communicated to other agents, may affect the emotional state of these agents either directly or indirectly. Regarding this effect we are left with hypotheses at the moment. These could be tested in computer simulations to investigate their impact on the possible emergence of a collective emotion – as it is done in the following. But there should be also the possibility to empirically test how individuals are affected by different emotional content, as discussed e.g. in [24].

With respect to the valence, i.e. the good or bad feeling, we have to take into account that there are two different kind of emotions in the system, positive ones represented by  $h_+(t)$  and negative ones represented by  $h_-(t)$ . Dependent on its own emotional state, an agent may be affected by these information in a different way. If we for example assume that agents with negative (positive) valence mostly respond to negative (positive) emotional content, we have to specify:

$$\mathcal{F}_v \propto \frac{r_i}{2} \{ (1 + r_i) f[h_+(t)] - (1 - r_i) f[h_-(t)] \} \quad (8)$$

where  $r_i(t) = \text{sign}(v_i(t))$  and  $f(h_{\pm}(t))$  are some functions depending either on  $h_+$  or on  $h_-$  only. Equation (8) then results in  $\mathcal{F}_v \propto f[h_-(t)]$  if the agents have a positive valence ( $r_i(t) = +1$ ), and in  $\mathcal{F}_v \propto f[h_+(t)]$  in the case of negative valence ( $r_i(t) = -1$ ).

If, on the other hand, there is evidence that agents, independent of their valence, always pay attention to the

prevalence of positive or negative emotional content, we may assume:

$$\mathcal{F}_v \propto g[h_+(t) - h_-(t)] \quad (9)$$

where  $g$  is some function of the difference between the two information available. Other combinations, for example agents with positive (negative) valence pay more attention to negative (positive) emotional content, can be tested as well. Some studies in psycho-physiology [25] provide initial results for heterogeneous emotional attention processes related to valence and arousal.

In the following, we may assume the case of equation (8), i.e. the valence increases with the respective information perceived by the agent. The impact, however, should depend also on the emotional state of the agents in a nonlinear manner. I.e. if an agent is happy (sad), it may become happier (more sad) if receiving information about happy (sad) agents or events, in a nonlinear manner, expressed in the general form:

$$f[h_{\pm}(t), v_i(t)] = h_{\pm}(t) \sum_{k=0}^n b_k v^k(t). \quad (10)$$

Here, it is assumed that the coefficients  $b_k$  are the same for positive and negative valences, which of course can be extended toward different coefficients.

## 2.6 Arousal and threshold

While the valence expresses the positivity or negativity of the emotion, the arousal measures the degree in which the emotion encourages or discourages activity. Only the latter is important for communicating the emotional content, which happens if a threshold  $\tau_i$  of the arousal is reached. Certainly, expressing the emotion should have some impact of the arousal, e.g. it is legitimate to assume that the arousal is lowered because of this action, or set back to the initial state in the most simple case. That means we should split the dynamics for the arousal into two parts, one applying before the threshold is reached, the other one when it is reached. For this, we redefine the arousal dynamics for  $a_i(t)$  given in equation (3) as the subthreshold dynamics  $\dot{\tilde{a}}_i(t)$  and set:

$$\dot{\tilde{a}}_i = \dot{\tilde{a}}_i(t) \Theta[\tau_i - a_i(t)] - a_i(t) \Theta[a_i(t) - \tau_i]. \quad (11)$$

As long as  $x = \tau_i - a_i(t) > 0$ ,  $\Theta[x] = 1$  and the arousal dynamics is given by  $\dot{\tilde{a}}_i(t)$ , equation (3), because  $\Theta[-x] = 0$ . However, after the threshold is reached,  $x \geq 0$ ,  $\Theta[x] = 0$  and  $\Theta[-x] = 1$ , i.e. the arousal is reset to zero.

It remains to specify the function  $\mathcal{F}_a$  for the subthreshold arousal dynamics. Since arousal measures an activity level, it would be reasonable to assume that agents respond to the sum of both positive and negative emotional content in a way that also depends on their own arousal in a nonlinear manner, regardless of the valence dimension. So, similar as for the valence, we may propose the

nonlinear dependence:

$$\mathcal{F}_a \propto [h_+(t) + h_-(t)] \sum_{k=0}^n d_k a^k(t). \quad (12)$$

Differently from the above assumption, we may argue that agents pay attention to the information only as long as their arousal is positive because negative arousals are associated with states of inactivities (tired, sleepy, depressed, bored). In this case, it is reasonable to assume e.g. that the impact of information increases linearly with the activity level:

$$\mathcal{F}_a \propto [h_+(t) + h_-(t)] a(t) \Theta[a(t)]. \quad (13)$$

To conclude the above description, we have set out a model where agents emotions are characterized by two variables, valence and arousal. These variables can be psychologically justified and most likely proxied empirically. The combination of these defines what kind of emotional content the agent expresses as an observable output. Again, this output is measurable and can be analysed. The way the emotional content is stored and distributed to other agents is explicitly modeled as part of a communication dynamics, which can be adjusted to specific practical situations.

## 3 Emergence of collective emotions

### 3.1 Valence dynamics

In our model, a collective emotional state can only emerge if a sufficient number of agents expresses their individual valences, which in turn depends on their arousal. The latter one gets above a critical threshold only if there is sufficient the emotional information  $h_+(t)$ ,  $h_-(t)$  available. However, this information is generated only by the agents. Hence, there is a circular causality between  $h(t) = h_+(t) + h_-(t)$  and  $a_i(t)$ .

In order to get a first insight into the dynamics, let us assume that there exist two different regimes: (i) a ‘silent’ regime where no sufficient emotional information is available, i.e.  $h(t) \rightarrow 0$ , and (ii) an ‘excited’ regime, where  $h(t)$  becomes large enough to affect enough agents. To simplify the study, we also assume that each agent is mostly affected by the information that corresponds to its valence state, as given by equation (8).

Then, neglecting any sort of random influences, the dynamics of the valence is expressed by:

$$\dot{v} = -\gamma_v v(t) + h_{\pm}(t) \{b_0 + b_1 v(t) + b_2 v^2(t) + b_3 v^3(t) + \dots\} \quad (14)$$

The stationary solutions for the valence then follow from the cubic equation:

$$v^3 + v^2 \{b_2/b_3\} + v \{(b_1 - \gamma_v/h_{\pm})/b_3\} + \{b_0/b_3\} = 0. \quad (15)$$

This allows to discuss the following cases:

- In order to allow for a solution  $v \rightarrow 0$  as requested,  $b_0$  should tend to zero as well, so we use  $b_0 = 0$  here. This leads to

$$v [v^2 + v\{b_2/b_3\} + \{(b_1 - \gamma_v/h_{\pm})/b_3\}] = 0. \quad (16)$$

- If positive and negative valences are treated as ‘equal’, there should be no ab initio bias towards one of them, which implies  $b_2 = 0$ . This gives, in addition to  $v = 0$ , the following two solutions:

$$v^2 = \frac{b_1 - \gamma_v/h_{\pm}}{b_3}. \quad (17)$$

These two solutions become real only if  $b_1 > \gamma_v/h_{\pm}$ . In this case, we have two equilibrium states for the valences which are symmetrical with respect to zero. Otherwise,  $v = 0$  is the only possible real solution.

So, dependent on the value of the information field  $h_{\pm}$  we can expect the two regimes: (i) the *silent regime* with  $h_{\pm} \rightarrow 0$  and  $v = 0$ , and (ii) the *excited regime* with the emergence of two different emotions, each of them centered around  $\pm b_1/b_3$  (provided the field is large enough). We note that these solutions are symmetrical, which can be changed by considering (a) a bias in the response ( $b_2 \neq 0$ ) or (b) differences in the two informations  $h_{\pm}$  (e.g. via different decay rates). It remains to be discussed whether a *coexistence* between the two collective emotional states is possible or the *prevalence* of one of them results. This leads us to the question of path dependence and emotional feedback of Section 3.2.

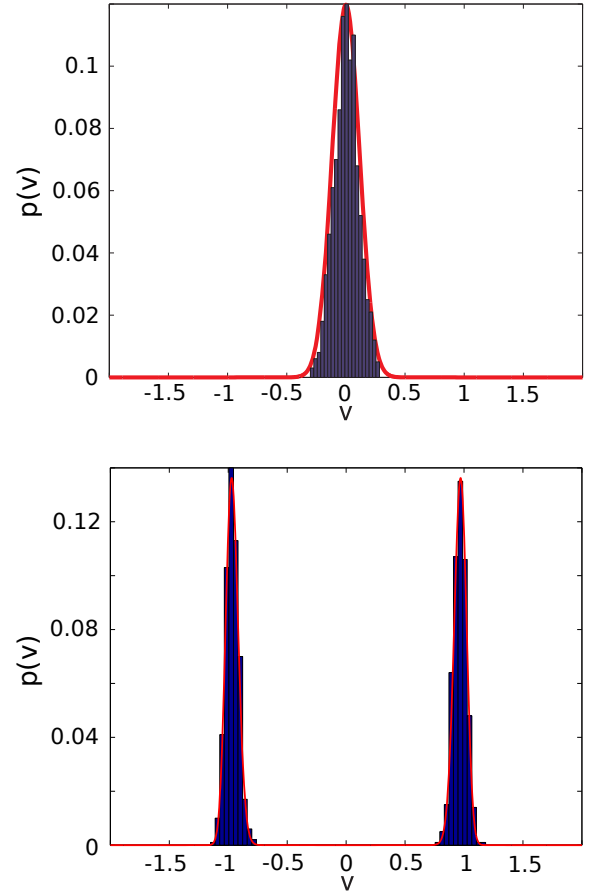
If, in addition to the deterministic dynamics specified above, we further consider stochastic influences as specified in the Langevin dynamics, equation (2), we can write up a dynamics for the valence distribution  $p(v, t)$ . Using equations (2), (15), this is given by the following Fokker-Planck equation:

$$\begin{aligned} \partial_t p(v, t) = & -\partial_v [(b_1 h_{\pm} - \gamma_v)v - b_3 h_{\pm} v^3] p(v, t) \\ & + \frac{A_v^2}{2} \partial_v^2 p(v, t). \end{aligned} \quad (18)$$

The stationary solution of the Fokker-Planck equation,  $\partial_t p(v, t) = 0$ , reads as:

$$p(v) = \frac{1}{\mathcal{N}_v} \exp \left\{ \frac{v^2(b_1 h_{\pm} - \gamma_v) - v^4(b_3 h_{\pm}/2)}{A_v^2} \right\} \quad (19)$$

$\mathcal{N}_v$  is the normalization constant resulting from  $\int_{-\infty}^{\infty} p(v) dv = 1$ . In accordance with the discussion above, the stationary valence distribution  $p(v)$  is *unimodal* with the maximum at  $v = 0$  if  $b_1 < \gamma_v/h_{\pm}$  and *bimodal* with the maxima given by equation (17) if  $b_1 > \gamma_v/h_{\pm}$ . In both cases, the variance of the distribution is determined by the strength of the stochastic force,  $A_v^2$ . This is shown in Figure 2 for two different values of  $h$ . The histograms result from computer simulations of the stochastic valence dynamics of 1000 agents for a given  $h$ , whereas the solid curves are given by the analytical solution of equation (19).



**Fig. 2.** (Color online) Analytical prediction of the valence distribution from equation (19) and histogram of valences for  $h = 0.25$  (top) and  $h = 17.5$  (bottom) with parameters  $N = 1000$ ,  $A_v = 0.15$ ,  $\gamma_v = 0.8$ ,  $b_1 = 1$  and  $b_3 = -1$ .

To conclude, this analysis has provided us with conditions regarding the influence of emotional information and the response to it, which may lead to the emergence of a collective emotional state. These conditions can be seen as testable hypotheses about the feedback between emotional information and individuals. If they hold true, we are able to predict the valence of a collective emotional state – which can then be compared to empirical findings. Deviations from these findings, on the other hand, allow us to successively refine the modeling assumptions made. Thus, the framework provided is a useful step toward a thorough understanding of collective emotions.

### 3.2 Arousal dynamics

In the previous section, we have detected an excited regime with a non-trivial valence, based on the assumption that the emotional information  $h(t)$  is large enough to affect the agents. The generation of such information, however, depends on the arousal dynamics. Specifically, in our model the arousal needs to reach a certain threshold  $\tau_i$  at which the agent expresses its emotion. In Section 2.6,

we have already introduced an arousal dynamics, equation (11), which distinguishes between a subthreshold dynamics, equation (3), and a dynamics at the threshold. By using the nonlinear assumption of equation (12) up to second order and neglecting stochastic influences for the moment, we get for the subthreshold regime (omitting the index  $i$  for the moment):

$$\dot{a} = -\gamma_a a(t) + h(t) \{d_0 + d_1 a(t) + d_2 a^2(t) + \dots\} \quad (20)$$

where  $h(t) = h_+(t) + h_-(t)$ . The stationary solutions follow from:

$$a^2 + a\{(d_1 - \gamma_a/h)/d_2\} + \{d_0/d_2\} = 0 \quad (21)$$

which allows to discuss the following cases. If we consider only the constant influence of  $h$ , i.e.  $d_0 \neq 0$ ,  $d_1 = d_2 = 0$ , or a linear increase with  $a$ , i.e.  $d_0 \neq 0$ ,  $d_1 \neq 0$ ,  $d_2 = 0$ , we arrive at only *one* stationary solution for the arousal, which depends on  $h$ :

$$a(h) = \frac{hd_0}{\gamma_a - d_1 h}. \quad (22)$$

It means that the agents tend to be always in an ‘excited’ regime, the level of which is determined by  $h$ . If it happens that the arousal reaches a value above the threshold,  $a(h) \geq \tau$ , then  $a(t)$  is arbitrarily set back to zero and then starts to reach  $a(h)$ , again. The proposed ‘silent’ regime would then be reached only if  $h \rightarrow 0$ .

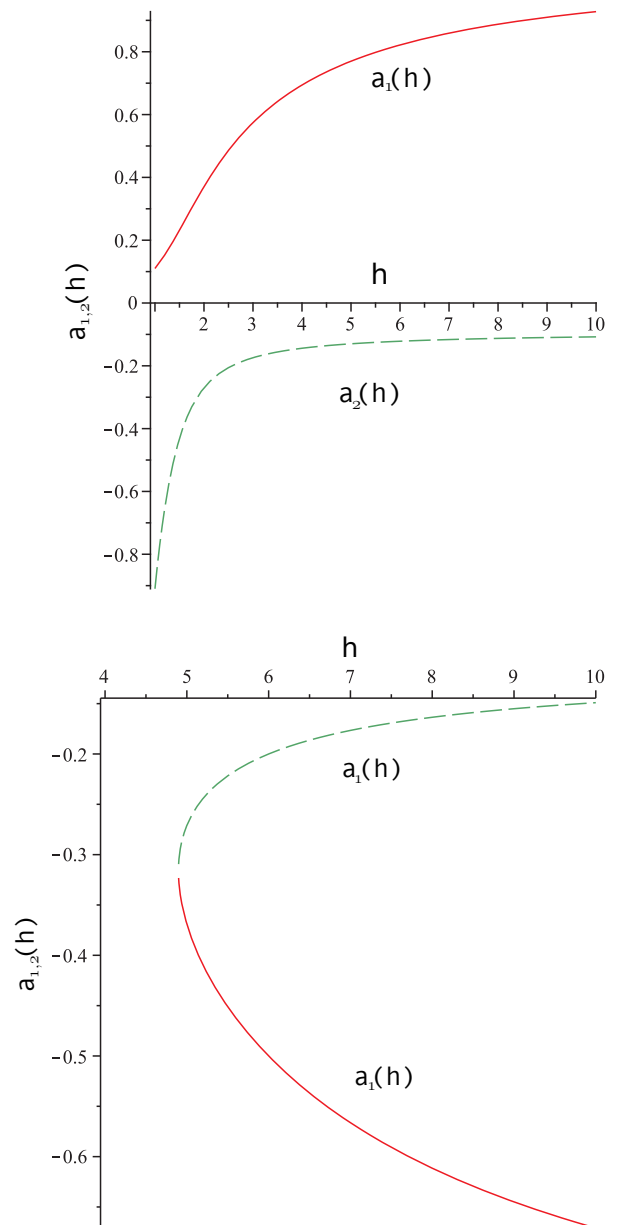
In order to allow for a dynamics where agents can stay at low values of the arousal even if  $h$  is large, we have to consider a non-linear influence of the emotional information  $h$ , i.e.  $d_0 \neq 0$ ,  $d_1 \neq 0$ ,  $d_2 \neq 0$  in the most simple case. Equation (21) then has two solutions

$$a_{1,2}(h) = \frac{1}{2} \left( \frac{\gamma_a/h - d_1}{d_2} \right) \pm \sqrt{\frac{1}{4} \left( \frac{d_1 - \gamma_a/h}{d_2} \right)^2 - \frac{d_0}{d_2}} \quad (23)$$

which are real only if

$$\left( \frac{d_1 - \gamma_a/h}{2d_2} \right)^2 > \frac{d_0}{d_2}. \quad (24)$$

From this restriction, we can infer some important conditions on the arousal dynamics. Provided  $d_0 > 0$ , inequality (24) is always fulfilled if  $d_2 < 0$  which, for a given  $h$ , implies a saturation in the feedback of the arousal on the arousal dynamics. Then we always have two real stationary solutions for the arousal, a positive and a negative one, shown in Figure 3. While the positive solution is stable for all values of  $h$ , the negative one is always unstable, as verified by the second derivative. This allows to infer the following dynamics for agents expressing their emotions: For agents starting with a small positive or negative arousal,  $a > a_2(h)$ ,  $a(t)$  may grow in time up to the stationary value  $a_1(h)$ , the level of which is determined by the emotional information available at that time. Only if  $a_1(h) > \tau_i$ , the agent expresses its emotions, which consequently sets back  $a_i(t)$  to zero, otherwise it remains at this



**Fig. 3.** (Color online)  $a_{1,2}(h)$  for the parameter set (31). There is no bifurcation present when  $d_2 = -0.5$  (top) but it appears when  $d_2 = 0.5$  (bottom).

subcritical arousal level. On the other hand, if the agent, because of some fluctuations, reaches the unstable negative arousal level  $a_2(h)$ , the feedback of equation (20) will further amplify the negative arousal to  $-\infty$ . This means that the agent never again expresses its emotions and ‘drops out’. If this happens to many agents, a collective emotion cannot be sustained. Which of the two cases is reached, crucially depends on the fluctuation distribution. Looking at the example of Figure 3, we can verify that initial fluctuations (for  $a = 0$ ) should not reach the level of 0.1, in order to prevent a ‘dropout’ of the agents.



The scenario looks different if, instead of a saturated dynamics with  $d_2 < 0$ , we assume  $d_2 > 0$ , i.e. a super-linear growth in the arousal. Inequality (24) then defines the range of possible values of  $d_2$  to guarantee two real solutions. As one can verify, there are real solutions already for very small values of  $h$ . In the following, we only concentrate on the range of sufficiently large  $h$  as shown in Figure 3. Then, both real solutions are negative and the one closer to zero is the unstable solution, whereas the most negative solution is stable. For agents starting with a small positive or negative arousal,  $a > a_1(h)$ ,  $a$  may further grow independent on the value of  $h$  until it reaches the threshold  $\tau_i$ , at which the agent expresses its emotion, i.e. agents do not remain at a subcritical arousal level forever. If the arousal is set back to zero and because of fluctuations reaches negative values  $a < a_1(h)$ , it will become more negative, but is always bound by the negative stationary value  $a_2(h)$ . I.e., the agent never ‘drops out’ entirely. Instead, even with a negative arousal, it can always get back into an active regime dependent on the fluctuation distribution. Consequently, the ‘non-saturated’ case defines the scenario where we most likely expect the emergence of collective emotions, where agents regularly express their emotions. However, such a scenario can never be sustained in a purely deterministic dynamics. Instead, spontaneous fluctuations are essential, and our analysis already tells us the critical size of the fluctuations needed (determined by  $A_a$ ). As we can verify in Figure 3, this critical fluctuation level depends on the total information  $h$ , which is not unrealistic, because more (diverse) information is also associated with more ambivalence.

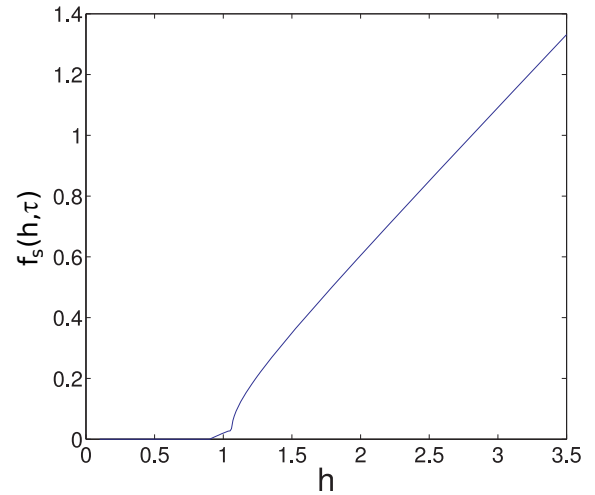
### 3.3 Expression of emotions

So far, we have identified critical regimes both in the valence and in the arousal dynamics, provided a given emotional information  $h$ . However, as explained above, this information is only generated by the agents above a critical arousal. Consequently, we need to ask what is the minimal time lapse before an agent reaches the threshold  $\tau$ , contributing to the emotional information. For simplicity, we take the delay time in equation (4) as  $\Delta t = 0$  for all expressions. The time lapse to reach the threshold is given by the dynamics of equation (20), which can be solved assuming a given value of  $h$ :

$$\int_0^T dt = T = \int_0^\tau \frac{da}{hd_2a^2 + (hd_1 - \gamma_a)a + d_0h}. \quad (25)$$

This solution assumes that  $h$  already exist, either because of an external information, or because it is generated by other agents. Hence, it is an adiabatic approximation of the full dynamics, which assumes  $\dot{h} = 0$ , this way describing the response of a single agent to the existing (stationary) field. The solution of equation (25) depends on whether the value of  $R(h) = 4d_2d_0 - (d_1 - \gamma_a/h)^2$  is positive or negative.

Following the discussion in the previous section, we now have to consider two different regimes for arousal dynamics, the saturated one ( $d_2 < 0$ ) and the superlinear



**Fig. 4.** (Color online) Frequency of emotional expression,  $f_s(h, \tau) = 1/T(h, \tau)$  for  $\tau = 0.5$  and the parameters (31). Below a critical value of  $h$ , an agent with threshold  $\tau$  would not express its emotions, but above the frequency of expression grows with the field  $h$ .

( $d_2 > 0$ ). In the saturated regime, always  $R(h) < 0$  and the solution is given by:

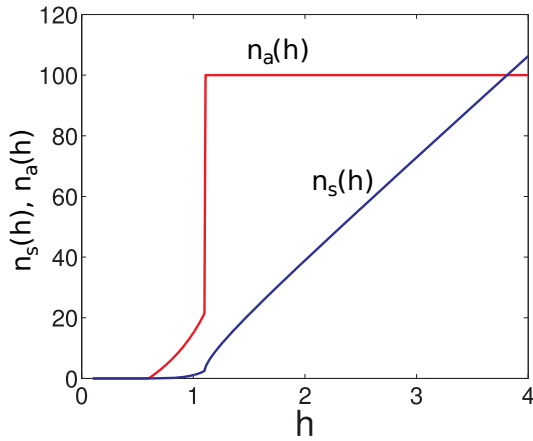
$$T(h, \tau) = \frac{2}{h\sqrt{-R(h)}} \operatorname{arctanh} \left( \frac{\sqrt{-R(h)}}{2d_0/\tau + d_1 - \gamma_a/h} \right). \quad (26)$$

For  $d_2 > 0$  we can have both  $R(h) < 0$  and  $R(h) > 0$  dependent on the choice of the other parameters. For  $R(h) > 0$ , the solution of equation (25) is given by:

$$T(h, \tau) = \frac{2}{h\sqrt{R(h)}} \operatorname{arctan} \left( \frac{\sqrt{R(h)}}{2d_0/\tau + d_1 - \gamma_a/h} \right). \quad (27)$$

In the superlinear regime, we expect that the agent is likely to express its emotions more than once (dependent on the fluctuations). In this case,  $T(h, \tau)$  gives the (idealized) periodicity of expressing the emotion, i.e. the time after which the agent on average reaches the threshold  $\tau$  again, after it was set back to zero when expressing the emotion last time. Figure 4 shows the frequency  $f_s(h, \tau) = 1/T(h, \tau)$  at which an agent expresses its emotions, dependent on the (quasistationary) value of  $h$ . We note that there is a nonmonotonous increase, i.e. below a critical value of  $h = h^*$  the frequency is zero, i.e. we do not expect a collective emotional state where agents more than once express their emotions, whereas for  $h > h^*$ , agents may regularly contribute emotional information, which means a collective emotion is sustained.

Based on the frequency of expression, we are able to calculate the average number of expressions per time interval,  $n_s$ , as defined in equation (6). Assuming  $N$  agents with a threshold distribution  $P(\tau)$ , the number of agents with a given threshold  $\tau$  is  $N(\tau) = NP(\tau)$ , whereas the frequency  $f_s(h, \tau)$  defines how often such agents reach an arousal above the threshold, forcing them to express their



**Fig. 5.** (Color online)  $n_a(h)$  (red) and  $n_s(h)$  (blue) versus field  $h$  for the parameter set (31). Above a critical level of information  $h$ , agents start to participate in the conversation. Their number  $n_a(h)$  grows very fast until the whole community is involved.

emotions. Assuming a uniform threshold distribution for simplicity, we can calculate for a given  $h$ :

$$n_s = N \int f_s(h, \tau) P(\tau) d\tau = \frac{N}{\tau_{\max} - \tau_{\min}} \int_{\tau_{\min}}^{\tau_{\max}} \frac{d\tau}{T(h, \tau)} \quad (28)$$

where  $T(h, \tau)$  is given by equations (26), (27).  $n_s$  is plotted in Figure 5. Above a critical value of the field  $h^*$ , the number of expressions per time interval increases monotonously, with a noticeable knee at the point where all agents become involved. Obviously, for lower values of  $h$  not all agents reach an arousal above the threshold, which prevents them from expressing their emotions. But at a characteristic value  $\hat{h}$ , the field is large enough to bring all their arousals above the threshold.

Similar to equation (5), we can also calculate the number of agents expressing their emotions at any given time  $t$  as:

$$N_a(t) = \sum_i 1 - \Theta[-f_s(h, \tau)]. \quad (29)$$

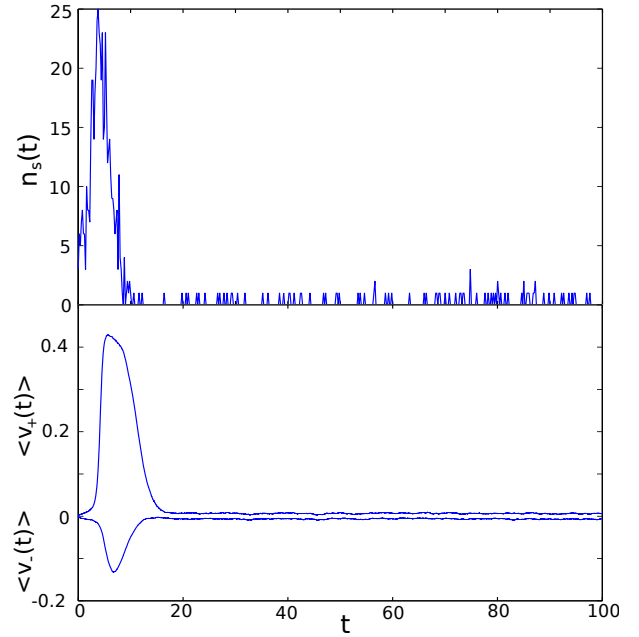
Again  $\Theta[x]$  is one only if  $x \geq 0$ , i.e. for agents with frequency zero the Heavyside function  $\Theta[-x]$  returns one. The *average* number of agents expressing their emotions per time interval is then, similar to equations (6), (28):

$$n_a = \frac{1}{t_{\text{end}}} \int_0^{t_{\text{end}}} N_a(t) dt = N \int (1 - \Theta[-f_s(h, \tau)]) P(\tau) d\tau \quad (30)$$

which can be calculated similarly to equation (28).  $n_a$  is plotted in Figure 5 as well, and one clearly identifies the critical  $\hat{h}$ , to involve all agents.

## 4 Computer simulations of collective emotions

Based on the analytical insights obtained, we eventually present the results of agent-based computer simulations.



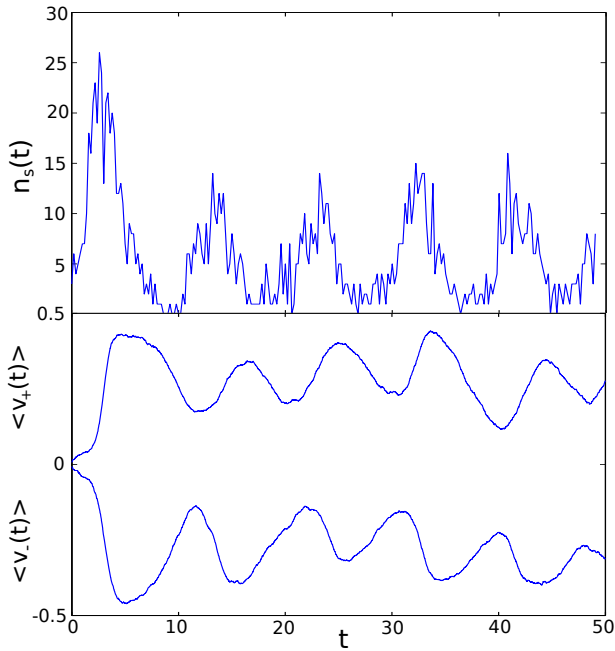
**Fig. 6.** (Color online) Agent expressions (top) binned with  $\delta t = 0.2$  and average positive and negative valences (bottom) and for a simulation with parameters (31) and  $d_2 = -0.1$ . A collective emotional state appears and disappears after some time, but never reappears again.

This means that we've implemented the individual dynamics given by the stochastic equations (2), (3) of  $N$  agents with a heterogeneous threshold distribution  $P(\tau)$ . The latter one is important as the process of forming a collective emotion needs generating emotional information,  $h(t)$ . There could be two possibilities to start this process: (i) an external trigger, expressed by  $I_{\pm}(t)$  in equation (7), (ii) initial fluctuations in the arousal which have to be large enough to push some of the agents above the threshold. Very similar to the model of social activation [26], it then depends on the distribution of thresholds and the feedback dynamics whether more agents become involved. For our simulations, we have chosen the parameters for the valence and the arousal dynamics,  $b_k$ ,  $d_k$  in such a way that a supercritical feedback between the emotional information generated and the activity of the agents is guaranteed. Specifically, we have chosen:

$$\begin{aligned} \gamma_v &= 0.5, A_v = 0.3, b_1 = 1, b_3 = -1, \gamma_h = 0.7, \\ A_a &= 0.3, \gamma_a = 0.9, d_0 = 0.05, d_1 = 0.5, \tau_{\min} = 0.1, \\ \tau_{\max} &= 1.1, N = 100, s = 0.1, h_0 = 0. \end{aligned} \quad (31)$$

That means that, thanks to our analytical efforts, we are likely to expect a collective emotion where most agents express their emotions at least once. Our main focus is therefore on the two different scenarios expressed by the parameter  $d_2 \leq 0$ , which result from the saturated or the superlinear feedback of emotional information on the arousal dynamics.

In the saturated case,  $d_2 < 0$ , we expect that a collective emotion may appear, but not be sustained because

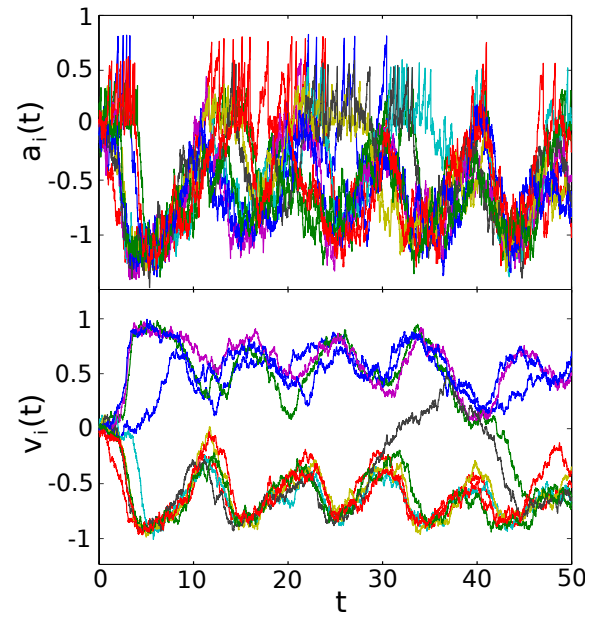


**Fig. 7.** (Color online) Agent expressions (top) binned with  $\delta t = 0.2$  and average positive and negative valences (bottom) and for a simulation with parameters (31) and  $d_2 = 0.5$ . A collective emotional state appears, fades out and reappears again.

agents have a tendency to ‘drop out’. This scenario is illustrated in Figure 6. Calculating the average number of expressions per time interval,  $n_s(t)$ , equation (6), we observe an initial burst of activity in the beginning, i.e. many agents contribute their emotional information, which then fades out, only keeping a random level of activity. That means, we observe indeed the emergence of a collective emotion, but this is not sustained because of the assumed saturation. This is also confirmed in Figure 6, which shows the averaged positive and negative valences of agents. We observe the emergence of a polarized state, where agents with strong positive and mildly negative emotions coexist, i.e. a bimodal valence distribution appears and remains for a while, before it disappears completely because agents ‘dropped out’. Consequently, the saturated regime allows the emergence of a collective emotion, but it is restricted to appear once and never again.

In the superlinear case,  $d_2 > 0$ , we expect the emergence of collective emotions more than once, i.e. they can fade out and be reestablished again. We consider this the more realistic scenario for applications to internet users, where the up and downs of collective emotions are indeed observed. Figure 7 illustrates this scenario in a way comparable to Figure 6. Here, we see waves of activity indicated by the number of emotional expressions per time interval. The respective averaged positive and negative valences also reflect these waves, i.e. we observe more or less polarized states dependent on the activity.

Consequently, the collective emotions not only emerge once, but are also sustained over a long period. The reason for this was already explained in Section 3.2. If agents have



**Fig. 8.** (Color online) Sample trajectories of the arousal (top) and the valence (bottom) of ten agents in a simulation with parameters as in Figure 7.

expressed their emotions and fall into a ‘careless’ state characterized by negative arousal, no new emotional information is produced. This in turn lowers the field  $h$ , which determines the stationary value of the negative arousal at which agents ‘rest’. The lower the field, the larger the stationary arousal, which eventually allows the fluctuations to push agents back into an active regime of  $a(t) > a_1(h)$ . To illustrate this, we have plotted the arousal of ten randomly chosen agents in Figure 8. The typical oscillatory behavior can be clearly seen. If the field is initially low, most likely  $a(t) > a_1(h)$ , i.e. agents arousal is increased until they express their emotions. This generates a high field. If agents arousal is set back to zero at high  $h$ ,  $a_1(h)$  is almost zero (as can be verified in Fig. 3), which means that most agents reach the stable stationary level of negative  $a_2(h)$ , at which they remain until  $h$  is lowered again.

The corresponding dynamics of the valence for the randomly chosen agents is also shown in Figure 8. One can notice a quite synchronized change of the emotions, which is not surprising as the dynamics mainly depends on the value of  $h$ , which is the same for all agents and all other parameters are kept constant. We can, of course, consider more heterogeneous parameters for the agents, to allow for more diversity. We wish to emphasize that agents do not always have the same emotion over time, some of the sample trajectories clearly show that agents switch from positive to negative emotions and vice versa.

To conclude, all simulations are consistent with the analytical results derived in the previous sections. Based on these results, we may be able to derive hypotheses about the behavior of emotional agents, which can be tested e.g. in psychological experiments.

## 5 Conclusion

The aim of our paper is to provide a general framework for studying the emergence of collective emotions in online communities. I.e. we are not particularly interested in the most complete description of individual emotions, but rather in an approach that allows to generate testable hypotheses about the conditions under which a particular collective dynamics can be observed. Nevertheless, we refrain from using ad hoc assumptions about the dynamics of individual agents which have been used in ‘sociophysics’ models of opinion dynamics, etc. Instead, our starting point is indeed a psychological theory of how individual emotional states should be described. Hence, our agent model is based on on *psychological variables* such as arousal  $a$  and valence  $v$ .

As the second important ingredient of our general framework, we explicitly address the *communication* between agents. This is very important to model online communities, where agents do not have a direct, face-to-face communication, but an indirect, time delayed communication, which is mediated by a medium. The latter stores the information expressed by the agents (mostly in terms of writings) and allows all agents to get access to this information at the same or a different times. Consequently, this medium provides a mean-field coupling between all agents, which is reflected by the so-called communication field  $h$  in our framework. By explicitly modeling the dynamics of the information stored, we consider the decentralized generation of emotional information of different types ( $h_+$ ,  $h_-$ ), at different times, the ‘aging’ of emotional information (i.e. the decrease in impact of older information) and the distribution of information among agents, for which also other than mean-field assumptions can be used.

As the third ingredient, we eventually model the impact of the emotional information on agents dependent on their emotional states. Here we assume a very general *nonlinear feedback* between the available information  $h$  and the individual valence  $v$  and arousal  $a$ . This allows to derive different hypotheses about the impact, which can be tested e.g. in psychological experiments. On the other hand, if such insight should become available to us, we are able to cope with these findings and to check their consequences on the emergence of collective emotions

It is a strength of our general framework that it allows an analytical treatment, to estimate the range of parameters under which the emergence of a collective emotion can be expected. In particular, we are able to specify the conditions for (a) a polarized collective emotional state (bimodal valence distribution), and (b) scenarios where agents express their emotions either once or consecutively (saturated vs. superlinear impact on the arousal dynamics). Such findings are important in order to later calibrate the model parameters against empirical data describing these different regimes.

Hence, our modeling approach offers a link to both psychological experiments with individuals, testing the hypotheses about the impact of emotional information, and to data analysis of emotional debates among in-

ternet users, determining model parameters for different scenarios.

Eventually, the general framework provided here is extensible and flexible enough to encompass different situations where collective emotions emerge. This is thanks to the distinction of the two different layers already depicted in Figure 1: the internal layer describing the agent and its emotional states and the external layer describing the communication process of expressing emotional information. If we, for example, want to apply this framework to emotions expressed in product reviews as e.g. analysed in [27], we have to consider that agents usually review a product only once, i.e. the saturated scenario for the arousal is more appropriate here. The dynamics for the valence, describing the emotional content, has to consider that the feeling of the agent also depend on the product quality  $q$  and the user preference  $u_i$ , i.e.  $v_i \propto |u_i - q|$ . Already such extensions are able to reproduce the distribution of the emotional content in reviews to a very remarkable degree [28]. Considering marketing campaigns in terms of external information versus a sole word-of-mouth spreading of emotional information also allows to capture different observed scenarios in generating emotional ratings [28].

In addition to these promising applications, there are other models which can be recasted in our general framework. For example, [29] or [30] have proposed agent-based models, where the information field does not feed back on the agent’s arousal, which is assumed as an exogenous constant probability of action. Consequently, instead of modeling the internal arousal dynamics explicitly as proposed in the general framework, these two models rather focus on the feedback between the expressed valence of different agents, which makes them special cases of the general framework.

To conclude, with our agent-based model we have provided a general framework to understand and to predict the emergence of collective emotions based on the interaction of agents with individual emotional states. As the framework is very tractable both in terms of mathematical analysis and computer simulations, we are now working on applying it to emotional debates observed in different online communities, such as in blogs, newsgroups, or discussion fora.

The research leading to these results has received funding from the European Community’s Seventh Framework Programme FP7-ICT-2008-3 under grant agreement No. 231323 (CYBER-EMOTIONS).

## References

1. J. Lorenz, S. Battiston, F. Schweitzer, Eur. Phys. J. B **71**, 441 (2009)
2. D. Bar-Tal, E. Halperin, J. de Rivera, J. Soc. Issues **63**, 441 (2007)
3. K.R. Scherer, T. Wranik, J. Sangsue, V. Tran, U. Scherer, Trends and developments: Research on Emotions **43**, 499 (2004)
4. J.A. Russell, J. Pers. Soc. Psychol. **39**, 1161 (1980)



5. L.F. Barrett, J.A. Russell, *Current Directions in Psychological Science* **8**, 10 (1999)
6. M. Yik, J. Russell, L.F. Barret, *J. Pers. Soc. Psychol.* **77**, 600 (1999)
7. *Appraisal processes in emotion: theory, methods, research*, edited by K.R. Scherer, A. Schorr, T. Johnstone (Oxford University Press, 2001)
8. J. Gratch, S. Marsella, *Cognitive Systems Research* **5**, 269 (2004)
9. J. Gratch, S. Marsella, N. Wang, B. Stankovic, *International Conference on Affective Computing and Intelligent Interaction* (2009)
10. C.M. de Melo, L. Zheng, J. Gratch, *Intelligent Virtual Agents*, **5773**, 301 (2009)
11. C.M. Jones, T. Troen, in *OZCHI '07: Proceedings of the 19th Australasian conference on Computer-Human Interaction* (ACM, 2007), p. 191, ISBN 978-1-59593-872-5
12. P. Sobkowicz, A. Sobkowicz, *Eur. Phys. J. B* **73**, 633 (2010)
13. R. Prabowo, M. Thelwall, *Journal of Informetrics* **3**, 143 (2009), ISSN 1751-1577
14. M. Thelwall, D. Wilkinson, S. Uppal, *J. Am. Soc. Inf. Sci. Technol.* **61**, 190 (2010)
15. F. Schweitzer, J. Holyst, *Eur. Phys. J. B* **15**, 723 (2000)
16. F. Schweitzer, *Brownian Agents and Active Particles. Collective Dynamics in the Natural and Social Sciences* (Springer, Berlin, 2003), with a foreword by J.D. Farmer
17. A. Flache, *Analyse & Kritik* **26**, 258 (2004)
18. P.S. Dodds, C.M. Danforth, *Journal of Happiness Studies*, **11**, 441 (2009)
19. M. Mitrović, B. Tadić, *Eur. Phys. J. B* **73**, 293 (2010)
20. A. Chmiel, J.A. Holyst, *Int. J. Mod. Phys. C* **21**, 593 (2010)
21. K. Sassenberg, M. Boos, *Group Process. Intergroup Relat.* **6**, 405 (2003)
22. P. DiMaggio, E. Hargittai, W.R. Neuman, J.P. Robinson, *Annu. Rev. Soc.* **27**, 307 (2001)
23. J. Adams, V.J. Roscigno, *Soc. Forces* **84**, 759 (2005)
24. L. Gianotti, P. Faber, M. Schuler, R. Pascual-Marqui, K. Kochi, D. Lehmann, *Brain Topogr.* **20**, 143 (2008)
25. M.M. Bradley, *Psychophysiology* **46**, 1 (2009)
26. M. Granovetter, *Am. J. Soc.* **83**, 1420 (1978)
27. J. Lorenz, *Eur. Phys. J. B* **71**, 251 (2009)
28. D. Garcia, F. Schweitzer, *Emotions in product reviews – Empirics and models*, to be submitted
29. S. Rank, *Docking Agent-based Simulation of Collective Emotion to Equation-based Models and Interactive Agents*, in *Proceedings of Agent-Directed Simulation Symposium, 2010 Spring Simulation Conference* (2010), pp. 82–89
30. A. Czaplicka, A. Chmiel, J.A. Holyst, *Acta Physica Polonica A* **117**, 688 (2010)