Supplementary Information to

Critical slowing down as early warning for the onset and termination of depression

van de Leemput et al. PNAS

Figures

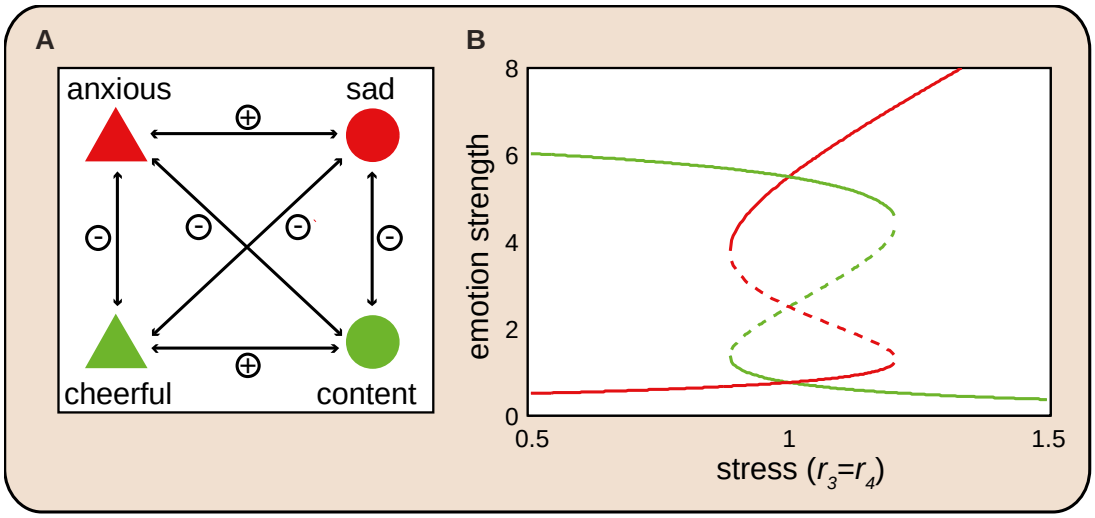

Fig. S1. The model. (**A**) A graphical representation of our simple dynamical model of four emotions. Emotions with the same valence have a positive effect on each other, while emotions of different valence have a strong negative effect on each other. (**B**) The stability properties of the deterministic part of the model (i.e. without noise) change if stress levels, represented by the growth rate of the two negative emotions (r_3 and r_4), change. Green lines represent positive emotions (x_1 and x_2), red lines represent negative emotions (x_3 and x_4). Solid lines represent stable states, and dashed lines unstable states. Far from the tipping point, at low stress levels, the network has only one stable state with high levels of positive emotions, and low levels of negative emotions. If stress levels increase, the network has two stable states: a 'normal state', and a 'depressed state', while at even higher stress levels, the system reaches a tipping point, at which the normal state disappears, and only one stable depressed state remains. Note that once the system is in the alternative depressed state, stress levels need to be decreased tremendously to trigger a backward shift.

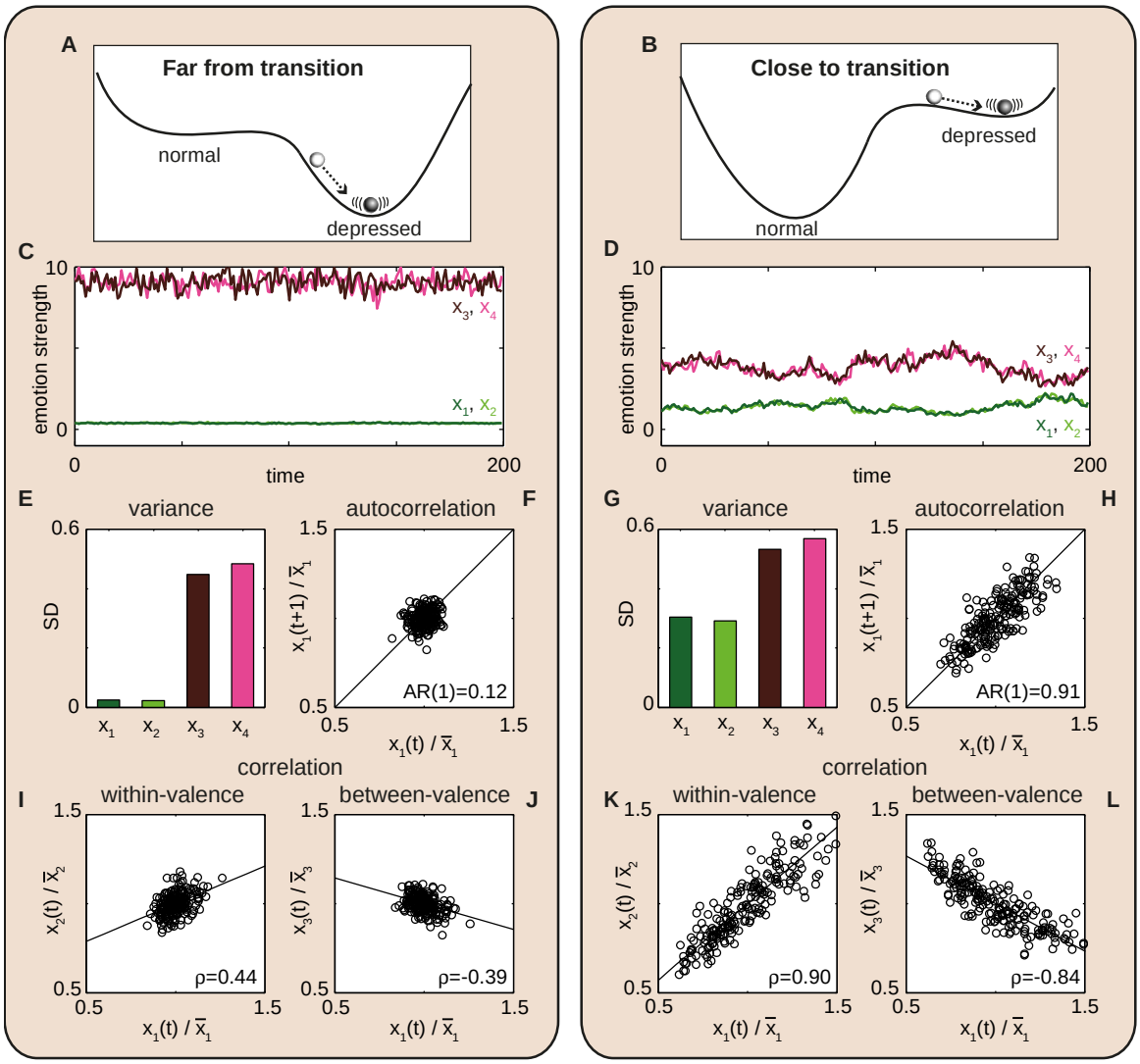

Fig. S2. Model simulations illustrating generic indicators of proximity to a tipping point from a depressed to normal state. Our model shows that the generic early warning signals that signal the proximity of a shift from a normal state towards a depressed state are also valid for the backward shift from a depressed state towards recovery. In that case, the stability of a depressed person may become more fragile close to the transition towards recovery (**B versus A**). Under a permanent regime of stochastic perturbations (**C and D**), slowing down near the tipping point results in higher variance (SD= standard deviation) (**G versus E**), higher temporal autocorrelation (AR(1)= lag-1 autoregression coefficient) (**H versus F**), and stronger correlation (ρ= Pearson correlation coefficient) between emotions with the same valence (**K versus I**), and between emotions with different valence (**L versus J**). Positive emotions are represented by x_1 and x_2, and negative emotions by x_3 and x_4. Parameters: left panels r_3=r_4=1.5, right panels r_3=r_4=0.9.

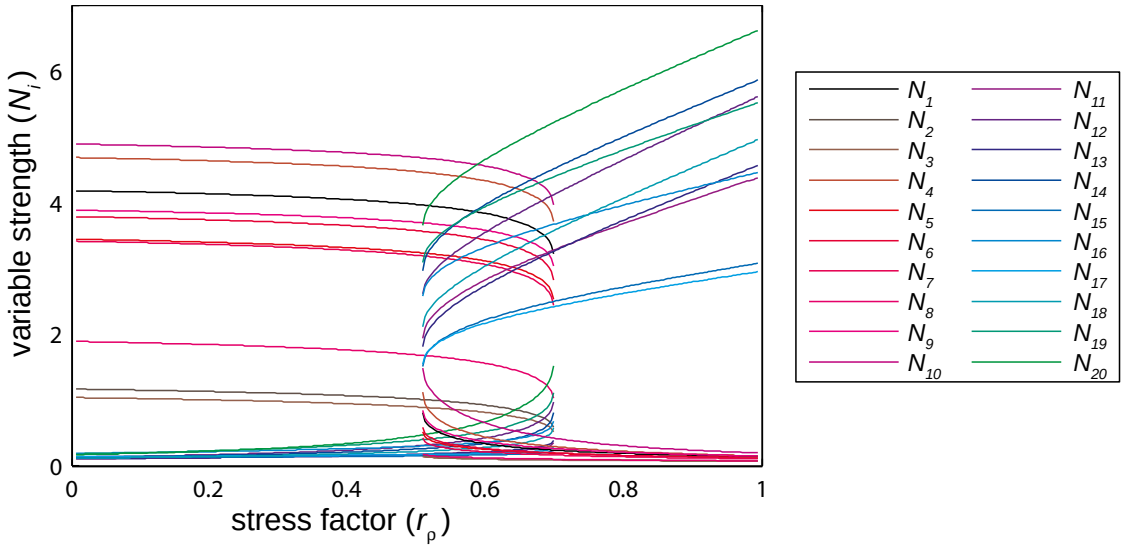

Fig. S3. Response of the network model to stress. The stability properties of the deterministic part of the model (i.e. without noise) change if stress levels, represented by r_ρ, change. Solid lines represent stable states, unstable states are not depicted. Far from the tipping point, at low stress levels, the network has only one stable state with one dominant cluster of network elements: the 'normal state'. If stress levels increase, the network has two stable states. Next to the 'normal state', another cluster can be dominant under the same conditions: the 'depressed state'. At even higher stress levels, the system reaches a tipping point, at which the normal state disappears, and only one stable depressed state remains.

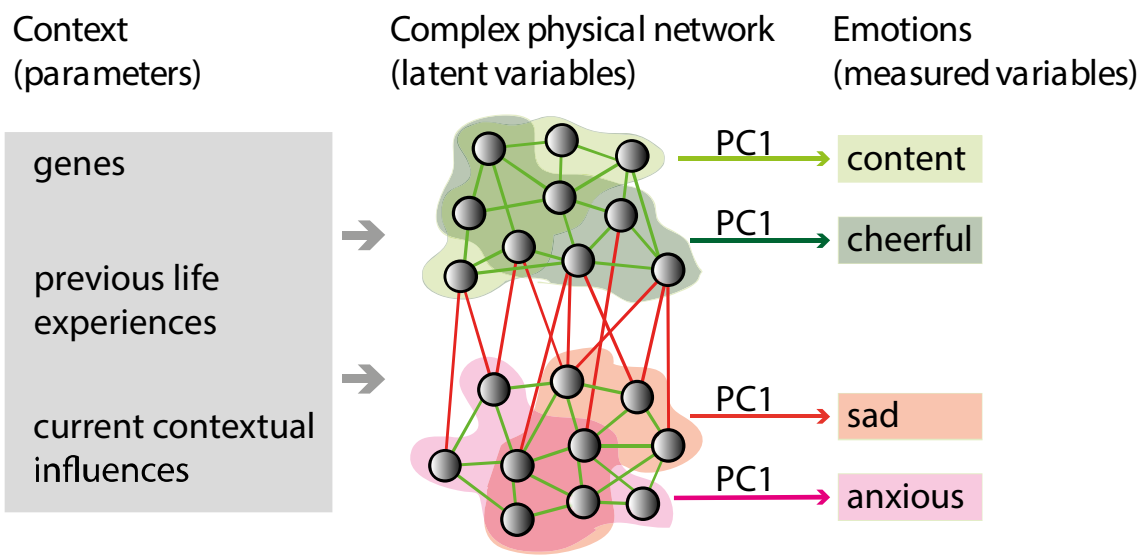

Figure S4. Illustration of the relation between the context, the complex physical network model (e.g. elements ranging from neurotransmitter and hormone concentrations to physical activity modes and social interactions) and the four newly defined variables. Note that the four variables are indirect indicators of parts of the complex system.

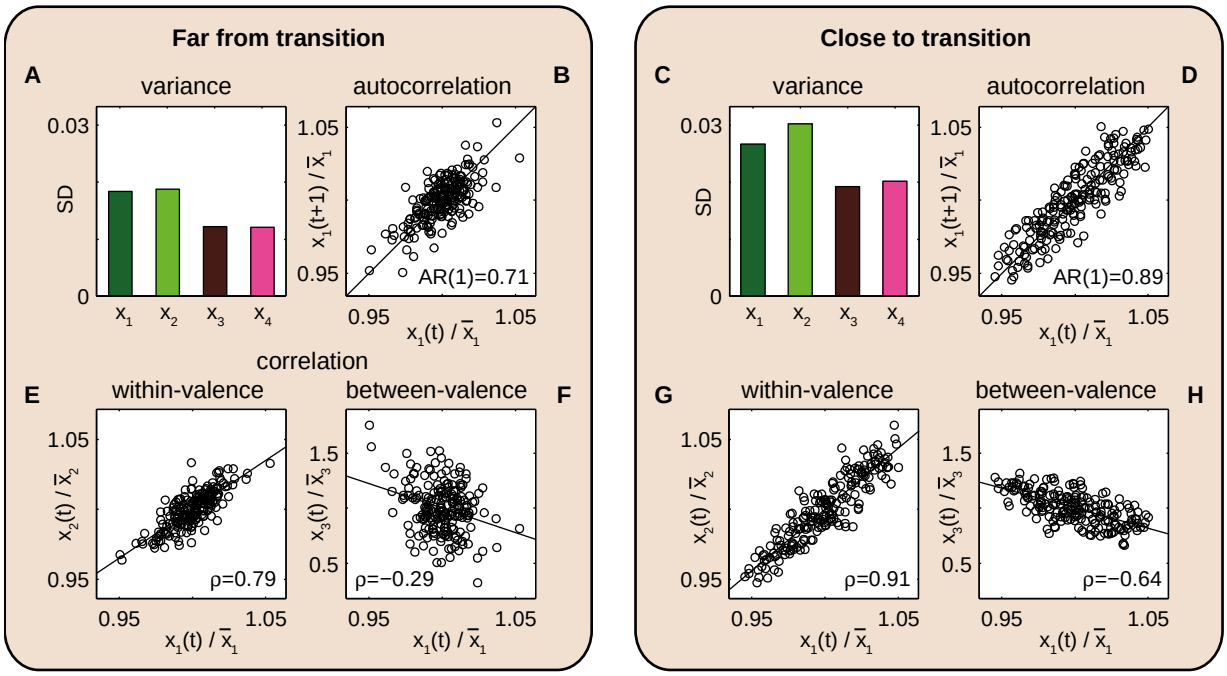

Fig. S5. Early warning signal analysis of model simulations of the four indirect indicators of the complex network. As for the four-component model with direct interactions, under a permanent regime of stochastic perturbations, slowing down near the tipping point results in higher variance (SD= standard deviation) (**A versus C**), higher temporal autocorrelation (AR(1)= lag-1 autoregression coefficient) (**B versus D**), and stronger correlation (ρ= Pearson correlation coefficient) between emotions with the same valence (**E versus G**), and between emotions with different valence **(F versus H**). Positive emotions are represented by x_1 and x_2, and negative emotions by x_3 and x_4. Parameters: left panels r_ρ=0.1, right panels r_ρ=0.68.

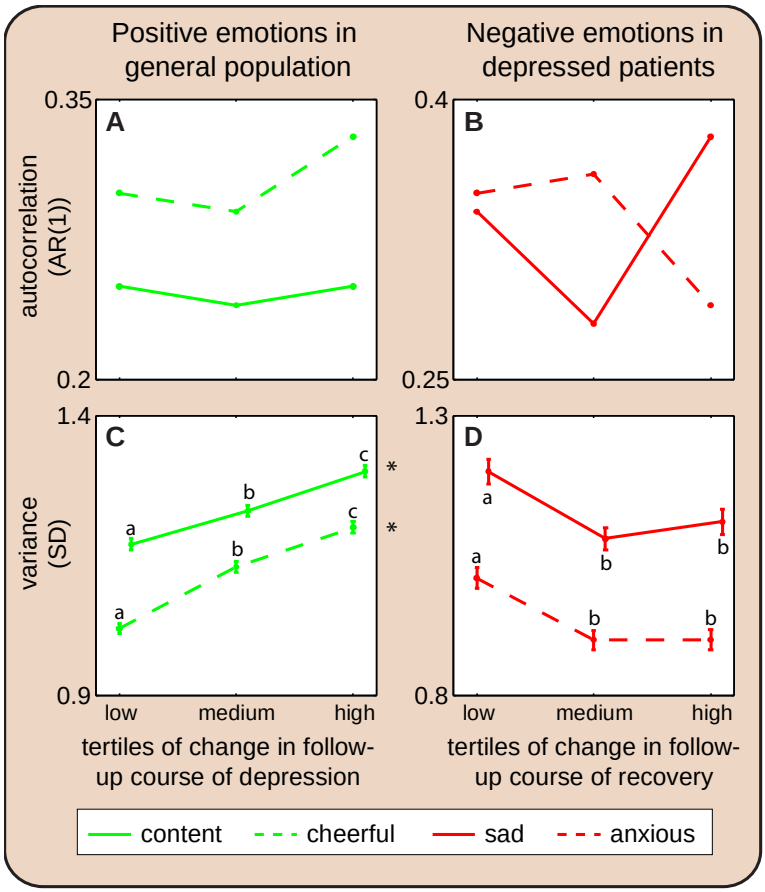

Fig. S6. Temporal autocorrelation and variance as a function of future symptoms. Increasing autocorrelation (AR(1) = mean lag-1 autoregression coefficient) (**A and B**) and variance (SD = mean standard deviation) (**C and D**) of positive emotions according to tertiles of development of future depressive symptoms in a general population (left panels), and of negative emotions according to tertiles of future recovery in depressed patients (right panels). For autocorrelation (**A and B**), we present data according to tertiles of change in follow-up course for illustrative purposes only, however, note that in the statistical analyses continuous variables were used. There are no significant trends in autocorrelation (positive interaction effect of future symptoms: $p<0.05$). For variance (**C and D**), error bars represent standard errors (SEs). Note that variance of negative emotions in the depressed population goes down with future recovery. This may be explained by differences in the mean (see Fig. S7). Asterisks indicate an overall significant upward trend in variance (overall tests: $p<0.05$). Mean values represented by different letters within emotions are significantly different (post-hoc tests: $p<0.05$).

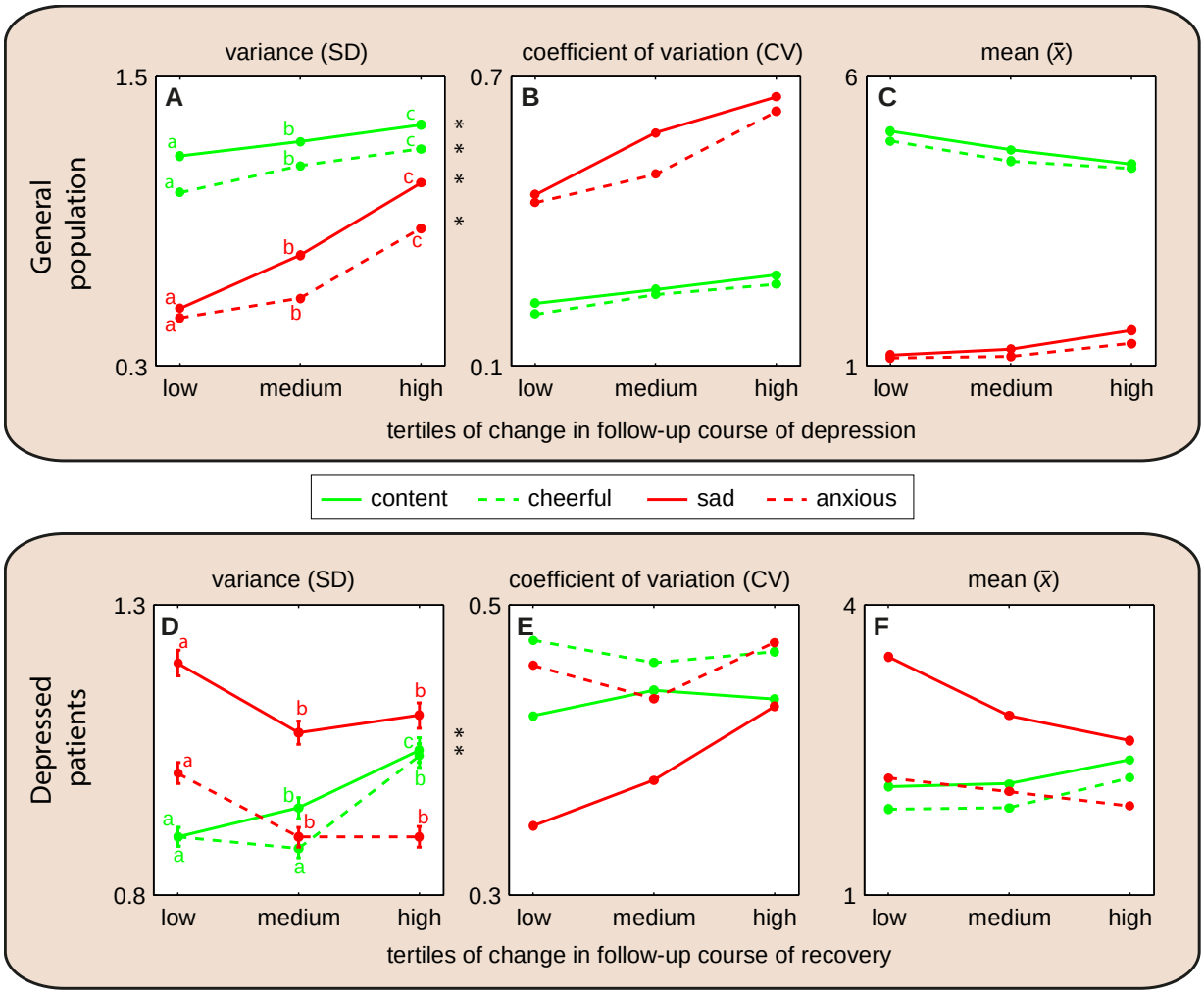

Fig. S7. The effect of critical slowing down on variance can be confounded by a change in the means. Variance (SD = mean standard deviation) (**A and D**), coefficient of variation (CV=SD/$\bar{x}$) (**B and E**), and mean affect level ($\bar{x}$) (**C and F**) according to tertiles of development of future depressive symptoms in a general population (n=535) (**upper panels**), and according to tertiles of future recovery in depressed patients (n=93) (**lower panels**). Note that for the general population, higher variance in individuals with higher future recovery is robust if corrected for the means, while for the depressed population, both higher variance of positive emotions, and lower variance of negative emotions, are not robust.

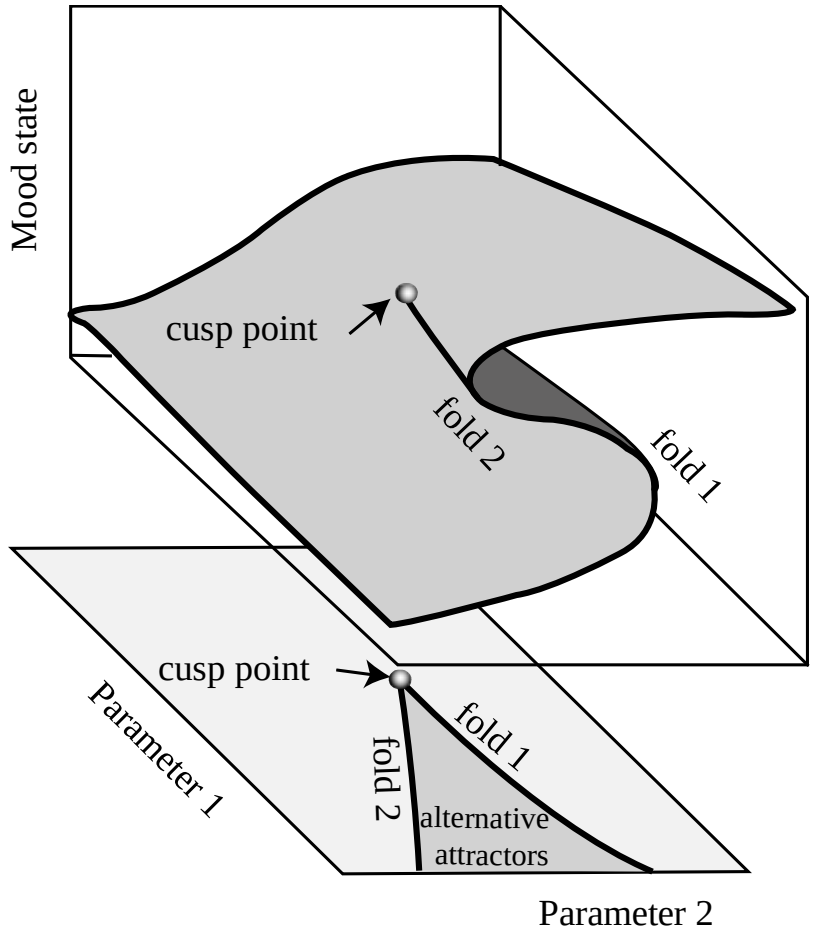

Fig. S8. The response of a dynamical system to a stressor (e.g. parameter 2) may be smooth or catastrophic depending on the strength of a positive feedback (e.g. parameter 1).The cusp point defines the parameter settings at which the system changes from smooth to catastrophic. The fold bifurcations define the parameter settings at which the system changes from two alternative stable states to one.

Tables

Table S1a. The socio-demographic and depression-related characteristics for the general population sample.

General population sample (n=535)			
	Mean (SD) or percentage	n (individuals) N (observations)	
Age	27.6 (7.8)	n=534	
Female gender	100%	n=535	
No/only primary school education	1%	n=4	
Secondary school education only	1%	n=6	
Intermediate vocational education	34%	n=184	
College/University	64%	n=341	
Baseline SCL-90-R (item average)	1.44 (0.51)	n=535	
Average follow-up SCL-90-R (item average)	1.47 (0.48)	n=535	
Baseline average rating (1-7) of *cheerful*	4.63 (0.86)	n=535	N=19,752
Baseline average rating (1-7) of *content*	4.77 (0.86)	n=535	N=19,660
Baseline average rating (1-7) of *anxious*	1.22 (0.38)	n=535	N=19,673
Baseline average rating (1-7) of *sad*	1.35 (0.52)	n=535	N=19,732
Average follow-up SCL-90-R per tertile (low, medium or high follow-up score)	low: 1.08 (0.06) n= 182	medium: 1.33 (0.09) n= 177	high: 2.02 (0.48) n=176
Baseline average rating (1-7) of *cheerful* per tertile of follow-up SCL-90-R score	4.90 (0.90)	4.54 (0.80)	4.43 (0.81)
Baseline average rating (1-7) of *content* per tertile of follow-up SCL-90-R score	5.07 (0.85)	4.73 (0.81)	4.51 (0.83)
Baseline average rating (1-7) of *anxious* per tertile of follow-up SCL-90-R score	1.13 (0.31)	1.16 (0.24)	1.38 (0.49)
Baseline average rating (1-7) of *sad* per tertile of follow-up SCL-90-R score	1.18 (0.43)	1.30 (0.41)	1.59 (0.62)

Table S1b. The socio-demographic and depression-related characteristics for the depressed patient sample.

Depressed patients (n=93)			
	Mean (SD) or percentage	n (individuals) N (observations)	
Age	41.7 (9.9)	n=93	
Female gender	40%	n=93	
No/only primary school education	19%	n=18	
Secondary school education only	27%	n=25	
Intermediate vocational education	39.8%	n=37	
College/University	10.8%	n=10	
Baseline HDRS-17 total score	24.0 (3.7)	n=93	
Follow-up HDRS-17 total score	12.5 (6.8)	n=93	
Baseline average rating (1-7) of *cheerful*	1.96 (0.92)	n=93	N=4.250
Baseline average rating (1-7) of *content*	2.19 (1.03)	n=93	N=4.270
Baseline average rating (1-7) of *anxious*	2.03 (1.40)	n=93	N=4.275
Baseline average rating (1-7) of *sad*	3.00 (1.32)	n=93	N=4.282
Intervention following baseline: **-combination of pharmacotherapy and supportive psychotherapy** **-imipramine (as part of a trial)** **-placebo (as part of a trial)**		n= 43 n=23 n=27	
Average follow-up HDRS-17 per tertile of change in follow-up HDRS-17 score (low, medium or high reduction in symptoms)	low: 19.1 (3.5) n= 33	medium: 12.2 (4.4) n= 32	high: 5.7 (3.4) n=28
Baseline average rating of *cheerful* per tertile of change in follow-up HDRS-17 score	1.87 (0.77)	1.90 (0.82)	2.15 (1.15)
Baseline average rating of *content* per tertile of change in follow-up HDRS-17 score	2.09 (0.92)	2.17 (0.94)	2.32 (1.24)
Baseline average rating of *anxious* per tertile of change in follow-up HDRS-17 score	2.17 (1.50)	1.97 (1.31)	1.93 (1.43)
Baseline average rating of *sad* per tertile of change in follow-up HDRS-17 score	3.51 (1.34)	2.79 (1.14)	2.62 (1.35)

Table S2. Regression analysis in which the interaction effect represents the extent to which autoregression coefficients increase with increased follow-up change in depressive symptoms.

Autocorrelation				
	General population		*Depressed patients*	
	Beta-coefficient of interaction effect size[α]	p-value	Beta-coefficient of interaction effect size[β]	p-value
Cheerful	0.014	0.537	0.008	0.017
Content	-0.007	0.738	0.006	0.100
Anxious	0.060	0.029	-0.002	0.662
Sad	0.065	0.024	0.005	0.135

α: follow-up average SCL-90-R depression score X 'emotion' moment (t-1) on 'emotion' moment (t)

β: decrease in HDRS-17 score from baseline to follow-up X 'emotion' moment (t-1) on 'emotion' moment (t)

Table S3a. The overall significance tests for differences between variances across the three tertile groups for the general population and the depressed patients.

Variance									
General population									
	Low FU symptoms		Medium FU symptoms		High FU symptoms		Overall Wald test		
	Coeff	SE	Coeff	SE	Coeff	SE	χ^2	df	p-value
Cheerful	1.02	0.009	1.13	0,01	1.20	0.010	165.52	2	<0.001
Content	1.17	0.010	1.23	0,01	1.30	0.010	68.13	2	<0.001
Anxious	0.50	0.004	0.58	0,005	0.87	0.008	1761.48	2	<0.001
Sad	0.54	0.005	0.76	0,007	1.06	0.009	2623.37	2	<0.001
Depressed patients									
	Low decrease in FU symptoms		Medium decrease in FU symptoms		High decrease in FU symptoms		Overall Wald test		
	Coeff	SE	Coeff	SE	Coeff	SE	χ^2	df	p-value
Cheerful	0.90	0.016	0.88	0.016	1.04	0.021	41.41	2	<0.001
Content	0.90	0.016	0.95	0.018	1.05	0.021	31.92	2	<0.001
Anxious	1.01	0.018	0.90	0.017	0.90	0.018	23.56	2	<0.001
Sad	1.20	0.022	1.08	0.020	1.11	0.022	17.16	2	<0.001

Table S3b. P-values of the post-hoc Wald tests for differences between variances across the three tertile groups for the general population and the depressed patients.

Variance			
General population			
	Low vs Medium FU symptoms	Low vs High FU symptoms	Medium vs High FU symptoms
Cheerful	<0.001	<0.001	<0.001
Content	<0.001	<0.001	<0.001
Anxious	<0.001	<0.001	<0.001
Sad	<0.001	<0.001	<0.001
Depressed patients			
	Low vs Medium decrease in FU symptoms	Low vs High decrease in FU symptoms	Medium vs High decrease in FU symptoms
Cheerful	0.337	<0.001	<0.001
Content	0.049	<0.001	<0.001
Anxious	<0.001	<0.001	0.883
Sad	<0.001	0.005	0.278

Table S4a. The overall significance tests for differences between correlations across the three tertile groups for the general population and the depressed patients.

Correlation									
General population									
	Low FU symptoms		Medium FU symptoms		High FU symptoms		Overall Wald test		
	Coeff	SE	Coeff	SE	Coeff	SE	χ^2	df	p-value
Anxious-sad	0.25	0.012	0.26	0.011	0.34	0.012	34.13	2	<0.002
Cheerful-content	0.50	0.009	0.54	0.009	0.56	0.009	22.19	2	<0.001
Anxious-cheerful	-0.16	0.012	-0.19	0.012	-0.21	0.012	10.20	2	0.006
Anxious-content	-0.19	0.012	-0.24	0.012	-0.28	0.012	26.54	2	<0.001
Sad-cheerful	-0.30	0.011	-0.35	0.011	-0.41	0.011	44.89	2	<0.001
Sad-content	-0.28	0.011	-0.34	0.011	-0.39	0.011	51.52	2	<0.001
Depressed patients									
	Low decrease in FU symptoms		Medium decrease in FU symptoms		High decrease in FU symptoms		Overall Wald test		
	Coeff	SE	Coeff	SE	Coeff	SE	χ^2	df	p-value
Anxious-sad	0.30	0.024	0.32	0.024	0.37	0.024	5.09	2	0.078
Cheerful-content	0.47	0.020	0.52	0.019	0.61	0.018	25.79	2	<0.001
Anxious-cheerful	-0.10	0.026	-0.12	0.026	-0.27	0.026	25.34	2	<0.001
Anxious-content	-0.14	0.026	-0.12	0.026	-0.22	0.027	8.19	2	0.017
Sad-cheerful	-0.30	0.024	-0.35	0.023	-0.43	0.023	16.82	2	<0.001
Sad-content	-0.31	0.023	-0.35	0.023	-0.36	0.025	2.20	2	0.332

Table S4b. P-values of the post-hoc Wald tests for differences between correlations across the three tertile groups for the general population and the depressed patients.

Correlation			
General population			
	Low vs Medium FU symptoms	Low vs High FU symptoms	Medium vs High FU symptoms
Anxious-sad	0.294	<0.001	<0.001
Cheerful-content	0.001	<0.001	0.225
Anxious-cheerful	0.107	0.001	0.112
Anxious-content	0.002	<0.001	0.032
Sad-cheerful	0.002	<0.001	<0.001
Sad-content	<0.001	<0.001	<0.001
Depressed patients			
	Low vs Medium decrease in FU symptoms	Low vs High decrease in FU symptoms	Medium vs High decrease in FU symptoms
Anxious-sad	0.478	0.027	0.129
Cheerful-content	0.075	<0.001	0.001
Anxious-cheerful	0.694	<0.001	<0.001
Anxious-content	0.659	0.024	0.007
Sad-cheerful	0.164	<0.001	0.008
Sad-content	0.249	0.168	0.787

Text

Text S1. Network model of latent variables

We developed a network model that serves as a hypothetical representation of the complex neurobiological system underlying the mood of an individual person. The network consists of twenty interacting latent variables. Each network variable represents one (unknown, but in principle measurable) component of the neurobiological system of that individual. Emotions are not represented directly as variables but are computed as principal components of simulation results of clusters of the network. In contrast with the simple model in the main text, they do not interact directly with each other. We demonstrate that such indirect indicators show the same behaviour in terms of early warning signals.

The network model was also based on the Lotka-Volterra model, describing the dynamics of interacting variables, representing the components of the neurobiological system:

$$\frac{dN_i}{dt} = r_i N_i + \sum_j^{20} C_{i,j} N_j N_i + \epsilon_N + \mu$$

where N_i represents the strength of network variable i, r_i represents the maximum rate of change of network variable i, C represents a matrix of interactions between network variables, μ represents a small continuous increase of the strength of a network variable (independent of their state) (μ=1), and ϵ_N is the stochastic part of the model represented by a Gaussian white noise process of mean zero and intensity σ^2/dt (σ=0.1) (i.e. additive noise).

We parameterized the network such that the system has two main clusters: network variables that are in the same cluster have a positive effect on each other, while variables of different clusters have a negative effect. The interaction strengths $C_{i,j}$, as well as the maximum rate of change (r_i), were randomly drawn from two uniform distributions. Positive interactions between network variables within a predefined cluster ranged from 0.003 to 0.005. Similarly, the negative interactions between variables of different clusters were drawn in a range between -0.002 and -0.004. The maximum relative rates of change (r_i) of the individual variables were assumed to be stress dependent, following:

$$r_i = r_{0,i} + r_\rho \rho_i$$

Maximum rates of change of network variables in a state without stress (r_0) are set to differ between the two clusters. In cluster 1 r_0 ranges from 0 to 1, while in cluster 2 r_0 ranges from 0 to 0.5. Stress is assumed

to influence the maximum rates by a factor r_ρ. Each network variable has a different sensitivity (ρ) to this stress factor. The sensitivity of variables in cluster 1 is assumed to be 0, while the sensitivity of variables in cluster 2 ranges from 0 to 1. For these parameter settings, this complex network has alternative stable states (Fig. S3).

In order to define four relevant indicators of dynamics in the network, we assume that each emotion is influenced by the dynamics of a subcluster of the network: each positive emotion is determined by seven of the ten variables of cluster 1, while each negative emotion is determined by seven of the ten variables of cluster 2 (Fig. S4). The subclusters that define the new variables contain overlapping network variables. Therefore, we simulated two time series with a different dominant cluster. We used each time series to perform two PCA analyses on seven variables of the dominant cluster. We used the first principal component (*PC1*) of each analysis to define the dynamics of the four new variables (x). For instance, the first variable (x_1) is defined as follows:

$$x_1 = \sum_j^7 PC1_j N_j$$

We simulated the dynamics of the complete model, and used the data of the four variables as input for the early warning signal analysis, as in the main text.

Importantly, in our network model, the four variables representing emotion strength (x) do not directly affect each other, they are simply indicators of the dynamics of a complex underlying network (Fig. S4). Our analyses show that the same early warning signals are expected if the variables are indirect indicators of a complex underlying system with tipping points between alternative stable state (Fig. S5). The predictions of critical slowing down are thus robust against this oversimplified way of representing emotions in the model of the main text.

Text S2. Supplementary methods

Inclusion criteria and final set of participants. Inclusion criteria in both studies were a DSM-IV diagnosis of major depressive disorder (MDD), age between 18 and 65 years, and a baseline score of ≥18 on the 17-item HDRS. Patients using psychotropic medications, other than low dose benzodiazepines, were excluded (1, 2). Of the 621 individuals of the general population sample, only 610 participated in ESM. Of this group 31 were excluded because of too few valid ESM measurements (3). Forty-four participants had missing data either at baseline or follow-up resulting in 535 individuals. In the depressed sample 118 were eligible to participate. Of those, six were excluded because of too few valid ESM measurements and 1 because of unavailability of emotion ratings in ESM. Additionally, 1 had missing baseline data and 17 had missing follow-up HDRS measurements. This resulted in a final sample of 93 participants.

Heteroscedasticity and normality. The current samples have 535 and 93 groups (individuals) with on average 37 and 45 observations, respectively, per individual. When checking our data, two main assumptions of the model did not hold for some of the analyses: homoscedasticity at level 1 (i.e., the variability of residuals within persons may differ from one person to the other) and normality (i.e., the distribution of scores within a person may not be normal). Violations of these assumptions were found through the inspection of residual plots. Estimates in the models may be slightly downwardly biased if the number of groups (level 2 units) is less than 50 and the normality assumption is violated. According to Hox (4) at least 50 level 2 groups (in this case individuals) are needed with 20 or more observations within each group in order to accurately estimate standard errors in case of violation of the normality assumption. Thus, according to Hox (4), the current sample sizes are adequate to yield accurate estimations of standard errors.

In order to test the potential influence of heteroscedasticity, all analyses were repeated with robust standard errors (using the so-called Huber–White or sandwich standard errors). These analyses yielded similar results and conclusions.

Estimating the potential function. We have considered the possibility to directly estimate the potential function. However, although the methodology is developed for a long time series (see e.g (5, 6)), the extension to our case is far from trivial. The reason is that the data consist of a sample of quite short time series, which do not yield enough information for estimating a person-specific potential function that is flexible enough (i.e., not restricted to a specific parametric form). In principle, this would be possible by setting up the estimation problem in the aforementioned multilevel modeling framework. However, this is a completely new methodology that has not been developed, let alone be sufficiently tested. Therefore, we have refrained in this paper from estimating the potential function.

Text S3. Individual and group responses

All people differ in their response to changing conditions and in their underlying emotional vulnerability. For each individual the dynamic interplay between emotions may differ. For example, some individuals quickly become anxious if something happens that makes them sad, while others don't have a strong connection between these two emotions (7). This may explain why some people slowly glide into a depression, while others shift much more suddenly and unexpectedly (Fig. S8). The result of the complex interplay between the multiple different emotional states people experience may thus differ from individual to individual and may impact on moment and timing of transition. We can hypothesize that the critical moment and speed with which a system may shift to another level of depressive symptoms is different per individual. When data of many different individuals are grouped together we expect –*at group level*- early warning signals to be associated with a dimensional change in depressive symptoms (since every system has its own point to shift), which is a reason for not categorizing by diagnosis status. This also illustrates a second reason: we do not necessarily expect that transition moments coincide with man-made arbitrary DSM-IV criteria. For some individuals critical shifts may occur at subclinical levels while for other individuals shifts occur to clinical levels of depression. As explained above each individual likely has his/her own mood set points and thresholds for tipping points, and some may even have no thresholds at all, but simply a smooth response to changing conditions. The results of the study support this view on transitions since indicators of critical slowing down predicted dimensional transitions towards higher or lower levels of depressive symptoms.

References

1. Peeters F, Nicolson NA, Berkhof J, Delespaul P, deVries M (2003) Effects of daily events on mood states in major depressive disorder. *J Abnorm Psychol* 112(2):203-211.
2. Barge-Schaapveld DQ, Nicolson NA, van der Hoop RG, De Vries MW (1995) Changes in daily life experience associated with clinical improvement in depression. *J Affect Disord* 34(2):139-154.
3. Delespaul P (1995) *Assessing schizophrenia in daily life: The experience sampling method* (University of Limburg, Maastricht)
4. Hox J (2010) *Multilevel analysis: techniques and applications (Quantitative Methodology Series)* (Routledge, New York) 2nd Ed
5. Brillinger DR (2007) Learning a potential function from a trajectory. *IEEE Signal Processing Letters* 14(12):1-4.
6. Wagenmakers E-J, Molenaar PCM, Grasman RPPP, Hartelman PaI, van der Maas HLJ (2005) Transformation invariant stochastic catastrophe theory. *Physica D* 211(3-4):263-276.
7. Wigman JTW, et al. (2013) Psychiatric diagnosis revisited: towards a system of staging and profiling combining nomothetic and idiographic parameters of momentary mental states. *PLoS ONE* 8(3):e59559-e59559.