A Dynamical Bifurcation Model of Bipolar Disorder Based on Learned Expectation

Shyr-Shea Chang¹ and Tom Chou¹,²

¹Department of Mathematics, University of California Los Angeles, Los Angeles, CA 90095, USA
²Department of Biomathematics, University of California Los Angeles, Los Angeles, CA 90095, USA

Keywords: bipolar disorder, behavioral approach system, dynamics, affective bias, prediction error

Abstract
Bipolar disorder is a common psychiatric dysfunction characterized by recurring episodes of mania and depression. Despite its prevalence, the causes and mechanisms of bipolar disorder remain largely unknown. Recently, theories focusing on the interaction between emotion and behavior, including those based on dysregulation of the so-called Behavioral Approach System (BAS), have gained popularity. Mathematical models built on this principle predict bistability in mood but do not involve intrinsic biological rhythms arising from interactions between mood and expectation. We develop and analyze a model with clinically meaningful and modifiable parameters that incorporates the interaction between mood and expectation and that exhibits a transition to limit cycle behavior when a mood sensitivity parameter exceeds a threshold value, signalling a transition to a bipolar state. The model also predicts that asymmetry in response to positive and negative events can induce unipolar depression/mania, consistent with clinical observations. Finally, we show how observed effects of lithium and antidepressant-induced mania can be explained within the framework of our proposed model.

Corresponding author: Tom Chou, tomchou@ucla.edu
INTRODUCTION

Bipolar disorder is characterized by cycling between manic and depressive episodes (Geller & Luby, 1997). Its prevalence is estimated to be 0.3%–1.5% of the total population (Weissman et al., 1996). The lifetime cost for a single patient can reach several million US dollars (Begley et al., 2001) and medication use associated with bipolar disorder comprises about 7% of that used to treat all mental disorders (Whiteford et al., 2013). Bipolar disorder has a serious societal impact, with early onset a major risk factor for suicide (Hawton, Sutton, Haw, Sinclair, & Harriss, 2005), and 65.5 work days lost per year per patient (Kessler et al., 2006). Despite the significance of bipolar disorder, there is limited understanding of its biological mechanism (Geddes & Miklowitz, 2013). Modern techniques such as functional magnetic resonance imaging (fMRI) have located neural circuits including limbic networks and attentional systems, whose dysfunction may be correlated with bipolar disorder (Chen, Suckling, Lennox, Ooi, & Bullmore, 2011; Strakowski, Adler, Holland, Mills, & DelBello, 2004). However, how the dysfunction of these circuits leads to emotional vulnerability remains unclear.

In order to understand the mechanism of bipolar disorder and accelerate the development of treatment (Geddes & Miklowitz, 2013) many mathematical models have been proposed and fit to experimental data. An oscillation in mood, either observed or self-reported, is the defining feature of bipolar disorder (Geller & Luby, 1997). Thus, early models focus on explaining this oscillation (Bonsall, Geddes, Goodwin, & Holmes, 2015; Daugherty et al., 2009; Goldbeter, 2011). The models describe mood as being formed from an intrinsically oscillatory brain circuit and explain self-reported mood scores as well as the effects of medication. Following these studies, a natural next step is to clarify the mechanism of the oscillations and distinguish key differences between normal individuals and patients with bipolar disorder (see also Goldbeter (2011)). A popular theory in psychology states that dysregulation of the Behavioral Approach System (BAS) and the resulting interaction between mood, expectation, and behavior can explain bipolar disorder (Urošević, Abramson, Harmon-Jones, & Alloy, 2008). Psychological observations provide evidence of malfunction of the BAS and models have been built to explain bipolar disorder based on this malfunction (Steinacher & Wright, 2013). A key difference between BAS-based models and previous models is that BAS models show bistability in mood instead of oscillations, and require external input or noise to trigger switching between states of mania and depression (Cochran, Schultz, McInnis, & Forger, 2017).
Can a model exhibiting periodic mood oscillations be derived by taking expectation and behavior into account? Recent psychological experiments have shown that emotion is affected by the mismatch between expectation and reality instead of the reward value (Rutledge, Skandali, Dayan, & Dolan, 2014). Based on these observations it has been proposed that the interaction between mood and expectation can explain bipolar disorder as well as depression (Eldar, Rutledge, Dolan, & Niv, 2016; Mason, Eldar, & Rutledge, 2017). While these studies give psychological insight on a promising mechanism for bipolar disorder, a formal mathematical analysis of this possibility is lacking.

In this work we develop and analyze a model inspired by Eldar et al. (2016) and Mason et al. (2017) that is based on an interaction between mood and expectation. The model incorporates psychologically measurable parameters which presumably can be tracked by experiments (for example see Pulcu and Browning (2017)). We prove that the model exhibits oscillatory mood behavior when a particular psychological parameter, the mood sensitivity, exceeds a threshold value. For mood sensitivities below the threshold, the system exhibits normal mood response. It has been reported (Gotlib, Krasnoperova, Yue, & Joormann, 2004; Pulcu & Browning, 2017) that normal subjects as well as patients with mental disorder respond differently to positive and negative events. Here, we also explore this possibility and show that extreme asymmetry in mood sensitivity can lead to bipolar disorder, while minor asymmetry can lead to depression and mania as suggested by previous work (Eldar et al., 2016). Finally, we model the effects of pharmaceutical intervention, including those of antidepressants and lithium. Our model predicts behavior that are consistent with previous models and clinical observations.

**MATHEMATICAL MODEL**

We propose a continuous time model based on interactions between the dynamical variables of mood $m(t)$, expectation $v(t)$, and reality $r(t)$:

$$\frac{dm}{dt} = \eta_m(fm + r - v) - km - k_3m^3$$

$$\frac{dv}{dt} = \eta_v(fm + r - v).$$

Here, $\eta_m$ and $\eta_v$ are learning rates for mood and expectation, respectively, $f$ is a scale factor for how mood responds to perceived reality, and $k, k_3$ are linear and cubic recovery rates for mood, respectively. The main mechanistic feature of the model is that positive and negative surprises, i.e. the difference
between perceived reality and expectation, drive mood in corresponding directions, which in turn adjust
the perceived reality and speed up the adaptation of expectation. In this sense the mood is analogous to
the momentum of a damped harmonic oscillator (Eldar et al., 2016). From our daily experience it is
apparent that mood changes the way we perceive reality: a minor drawback may have no effect on us
when we are happy, but can be a source of depression if we are not in good spirits. The term \( fm + r \) in
our model represents, in a linear way, the modification that mood has on reality. Thus \( fm + r - v \) reflects
the extent to which an individual is surprised and how strong he or she should respond. Unlike
expectation, mood returns to a baseline level, even after winning a lottery (Brickman, Coates, &
Janoff-Bulman, 1978). A recovery “force” for the mood is captured by the \(-km\) term, where \( k^{-1}\)
represents the time scale for this recovery. Finally, if mood is viewed as a physiological quantity, there
should be upper and lower bounds limiting its magnitude. This suggests a higher order recovery term in
our model since, as we will see in the Results section, mood can grow without bound if there is no higher
order recovery term. Therefore we add a term \(-k_3m^3\) (corresponding to a quartic “potential”) to limit the
mood amplitude.

The reality \( r(t) \) is derived from external events and is not affected by personal mood or expectation. This
assumption distinguishes the proposed model from those based on the Behavioral Approach System
(BAS) (Steinacher & Wright, 2013). The model can also be written in terms of a single nonlinear
oscillator in mood (assuming that \( r(t) \) is differentiable):

\[
\frac{d^2 m}{dt^2} - (f\eta_m - k - \eta_v - 3k_3m^2)\frac{dm}{dt} + \eta_v km + \eta_v k_3 m^3 = \eta_m \frac{dr}{dt}.
\]

(3)

This is a Liénard equation (Strogatz, 2014) similar to the van der Pol oscillator invoked in previous
theories (Bonsall et al., 2015; Daugherty et al., 2009). The main new features here are the forcing term
\( \eta_m \frac{dr}{dt} \) that depends on changes in reality, the higher order term \( \eta_v k_3 m^3 \), and a possibly non-constant
parameter \( \eta_m \) as we will explore later in this section.

Throughout this paper we will explore the effects of two forms of the reality function \( r(t) \): a constant
\( r(t) = r_0 \), and a random \( r(t) \). In the random case, we assume a piecewise constant form for \( r(t) \) with
normally distributed values and log-normally distributed times between jumps. This functional form
reflects the abrupt nature of changes in reality such as salary raises or the death of relatives cause a
dramatic change that lasts for certain period of time. We set the mean and standard deviation of \( r(t) \) to be 0, \( \sigma_r \). The time intervals between jumps in \( r(t) \) are drawn from a log-normal distributions with mean of log time \( 1/k_r \) and standard deviation of the log time \( 1/k_r \). The parameters \( f, k, k_3, \eta_v \) are treated as positive constants throughout the paper. It has also been shown that learning rates \( \eta_m \) can be different for positive and negative events (Pulcu & Browning, 2017), which we model using a Heaviside function of \( f_m + r - v \):

\[
\eta_m = \begin{cases} 
\eta_m^+ & \text{if } f_m + r - v > 0 \\
\eta_m^- & \text{if } f_m + r - v \leq 0
\end{cases}
\]  

(4)

where \( \eta_m^+; \eta_m^- \) are positive constants. We will show in the Results section how asymmetry in \( \eta_m \) (the case \( \eta_m^+ \neq \eta_m^- \)) can lead to mental disorders. The parameters are tuned such that the timescale of mood variation matches the experimental data in Bonsall et al. (2015). To better connect our results with clinical observations, we calculate QIDS-SR16 (Quick Inventory of Depressive Symptomatology) scores (Rush et al., 2006) from our model. The QIDS-SR16 (QIDS for short) is commonly used for analyzing and testing treatments of bipolar disorder (Bonsall, Wallace-Hadrill, Geddes, Goodwin, & Holmes, 2012; Holmes et al., 2016) and consists of a 16-item self test that measures the level of depression. We calculate this score by taking \(- \min(0, m)\) since negative mood corresponds to depression. The system (1,2) is solved by explicit 4th to 5th order Runge-Kutta solvers, carried out by the \texttt{ode45} function in MATLAB.

**RESULTS**

**Mood and expectation become more oscillatory as the mood sensitivity increases**

For normal subjects we expect that if the reality \( r(t) = r_0 \) is constant the expectation should approach \( r_0 \) and the mood will relax to zero as there is no additional stimuli; this justifies shifting \( r_0 \to 0 \) without loss of generality and linearizing Eqs. (1-2) around the fixed point \((m, v) = (0, 0)\). In this way we can define the parameter regime within which the origin becomes linearly unstable (suggesting the onset of bipolar disorder) and which parameters are crucial in this process. In this subsection we assume symmetry in \( \eta_m \), i.e. \( \eta_m^+ = \eta_m^- \), to gain insight of the model before we study more general cases.

Upon linearizing Eqs. (1-2) or Eq. 3 about \((m, v) = (0, 0)\) for \( r = 0 \), we find the two eigenvalues
\[ \lambda_\pm = \eta_m f - \eta_v - k \pm \frac{\sqrt{\Delta}}{2}, \]

where the discriminant

\[ \Delta \equiv (\eta_m f - \eta_v - k)^2 - 4\eta_v k. \]

Thus, the origin is linearly stable when \( f\eta_m - k - \eta_v < 0 \) and unstable when \( f\eta_m - k - \eta_v > 0 \). This analysis agrees with that of Eldar et al. (2016) in which bipolar disorder arises either when \( f \) is large or \( \eta_m \gg \eta_v \). Here, we focus on mood and base our study on the quantity \( f\eta_m \), which we call the mood sensitivity parameter. In the linearly stable case (i.e. normal individuals), the system can support transiently oscillating behavior in mood and expectation, similar to the dynamics of a damped harmonic oscillator (Marion, 2013). Different types of oscillatory behavior can explain the differences in e.g. normal and cyclothymic personality (Mason et al., 2017).

The frequency of the oscillation of the solution is characterized by the imaginary part of the eigenvalues, determined by the sign of \( \Delta \). When \( \Delta \) is positive there will be no oscillation in the solutions, while negative \( \Delta \) corresponds to oscillatory solutions, with oscillation frequency determined by \( \sqrt{|\Delta|}/2 \). As a function of the mood sensitivity parameter \( f\eta_m \), we see that \( \Delta \) is a parabola with minimum at \( f\eta_m = \eta_v + k \), the critical value for linear stability, with a negative discriminant \(-4\eta_v k\). Thus, as \( f\eta_m \) increases toward the critical value \( \eta_v + k \), the mood and expectation become oscillatory with the frequency in the oscillations increasing. As the mood sensitivity parameter \( f\eta_m \) exceeds the critical value, a Hopf bifurcation occurs, the linearized dynamics are unstable, and linear analysis can no longer predict system behavior. This argument suggests that the mood fluctuation of normal individuals increases as the mood sensitivity increases. We verify these arguments by simulating Eqs. (1,2) using constant \( r(t) = 0 \) and different mood sensitivity parameters. The numerical solutions show that the oscillation frequency in mood and expectation increases as the mood sensitivity \( f\eta_m \) becomes larger, as predicted by our linear analysis (Fig. 1A, B). In the case of random reality the linear analysis does not fully apply, as it only works with constant \( r(t) \). The numerical simulations show that for larger \( f\eta_m \), expectation \( v(t) \) deviates more from reality \( r(t) \) and that mood \( m(t) \) experience higher variations about the zero baseline (Fig. 1C, D). These results suggest that the mood sensitivity parameter controls a
Figure 1. The mood and expectation of normal subjects become more oscillatory as the mood sensitivity $f_\eta m$ increases towards the critical value $\eta v + k$ from below. (A) Oscillations in expectation are highly damped for normal subjects (blue solid, $f_\eta m = 0.3(\eta v + k)$), but become less damped when the mood sensitivity increases (green dotted, $f_\eta m = 0.6(\eta v + k)$, and red dash-dot, $f_\eta m = 0.9(\eta v + k)$). Since we start the solutions at $(m, v) = (0, -1)$, the constant reality $r(t > 0) = 0$ represents a permanent increase in reality from $r(t < 0) = -1$. The numerical values $\eta v = 0.37$, $f = 0.3$, $k = 0.37$, and $k_3 = 2.8 \times 10^{-3}$ are used in all figures. (B) The mood shows similar oscillatory behavior that become less damped with increasing mood sensitivity. (C) When subjected to random reality events individuals with large mood sensitivity exhibit larger responses in expectation. (D) Similarly, the fluctuation in mood is greater for individuals with larger mood sensitivity under random reality conditions. Realizations of the random reality function are generated as described in the Mathematical Model section, with $\sigma_r = 2$, $k_r = 1$. In (C) and (D), mood and expectation are initialized at $(m, v) = (0, 0)$.

spectrum of personality responses, from normal to cyclothymic, and is a key determinant in triggering bipolar disorder as its threshold is exceeded.

A limit cycle occurs as mood sensitivity crosses the critical value, representing a bipolar state

Once the mood sensitivity $f_\eta m$ exceeds the threshold $\eta v + k$, linear analysis no longer holds since the origin becomes unstable and nonlinearities become important. However, for two dimensional systems we can rely on the Poincaré-Bendixson theorem to predict the existence of a limit cycle, a periodic solution that attracts solutions starting nearby (Strogatz, 2014). For this analysis, and in the rest of this subsection, we still assume $\eta m^+ = \eta m^-$ and a constant $r(t) = 0$. Since the origin is linearly unstable, we search for a limit cycle by constructing an outer boundary on which the vector fields are pointing inward. One way of
finding this boundary is to draw a rectangle whose edges connect two nullclines \( v = (f - \frac{k}{\eta m})m - \frac{k_3}{\eta m}m^3 \)
and \( v = fm \). Since \( \eta^+_m = \eta^-_m \) both nullclines are rotationally symmetric, we can find the distance to the
right edge of the boundary by setting \(-fm^*\) equal to the \( m \)-nullcline:

\[-fm^* = (f - \frac{k}{\eta m})m^* - \frac{k_3}{\eta m}m^{*3}, \tag{7}\]

which leads to

\[m^* = \sqrt{\frac{2f\eta_m - k}{k_3}}. \tag{8}\]

Thus, the rectangle with vertices \((\pm m^*, \pm fm^*)\) serves as an outer boundary confining all trajectories that
start inside it, leading to existence of a limit cycle (Strogatz, 2014). This result, along with the instability
of the \((m, v) = (0, 0)\) state as \( f\eta_m \) surpasses \( \eta_v + k \), implies a supercritical Hopf bifurcation at

\[f\eta_m = \eta_v + k. \]

Psychologically, this means that the expectation and mood persistently oscillate under
constant reality conditions, in sharp contrast to the behavior in a normal nonbipolar state (Fig. 2A, B). It
is analytically difficult to predict how the amplitude of the oscillation scales with the psychological
parameters since this requires analytically solving the nonlinear system. However, the formula for outer
boundary Eq. (8) could give us a prediction. Eq. (8) predicts that, after the onset of bipolar disorder, the
mood sensitivity \( f\eta_m \) still positively correlates with the mood amplitude. This prediction is numerically
verified by simulations for large \( f\eta_m \) (Fig. 2C), suggesting that the mood sensitivity parameter plays an
essential role even after the onset of bipolar disorder. How the amplitude of the oscillations depends on
the mood sensitivity \( f\eta_m \gtrsim \eta_v + k \) can be derived using weakly nonlinear analysis (Bender & Orszag,
1999) of Eq. 3 but will not be treated here.

While the current analysis applies only in the case of constant reality, the qualitative behavior of
persistent oscillations do not change even if the reality \( r(t) \) varies in time. Simulations show that the
oscillations in mood and expectation are not destroyed by changes in reality but take on an autonomous
nature (Fig. 2D, E). The QIDS score for the bipolar case shows intermittent peaks that match
qualitatively with experimental data (Bonsall et al., 2015; Bopp et al., 2010). Together, our analyses and
simulations show the onset of bipolar disorder as the mood sensitivity parameter \( f\eta_m \) crosses a critical
value, leading to persistent oscillations in mood and expectation qualitatively similar to those observed in
mood profiles of bipolar patients.
Figure 2. Our theory predicts that the onset of bipolar disorder occurs through a supercritical Hopf bifurcation as the mood sensitivity $f_{\eta_m}$ crosses the threshold value $\eta_v + k$ and a limit cycle in mood $m(t)$ is established. (A) In a bipolar state the expectation $v(t)$ (dotted green) persistently oscillates, in contrast to normal subjects (solid blue). We set the reality $r(t > 0) = 0$ and use $(m, v) = (0, -1)$ as the initial condition. The bipolar state is modeled using $f_{\eta_m} = 1.5(\eta_v + k)$ while the normal state is computed using $f_{\eta_m} = 0.3(\eta_v + k)$. The numerical values $\eta_v = 0.37$, $f = 0.3$, $k = 0.37$, and $k_2 = 2.8 \times 10^{-3}$ are used in all plots. (B) Mood of bipolar subjects also persistently oscillates. (C) The magnitude of mood oscillations increases as the mood sensitivity $f_{\eta_m}$ increases. The amplitude of oscillations obtained from numerical simulations (green stars) are compared well to amplitude estimates using Eq. (8) (black dots) when $f_{\eta_m} \gg \eta_v + k$. (D) Expectation $v(t)$ in the bipolar state responds to changes in reality but remaining oscillatory (green dashed). This behavior is distinct from the expectation of normal subjects (solid blue curve) that more closely follow the reality function. (E) Under the same reality function as in (D), the mood is much more oscillatory in the bipolar state (green dashed curve) than in the normal state (solid blue curve). (F) The model predicts intermittent spikes in the QIDS score. Realizations of the reality function are generated as described in the Mathematical Model section, with $\sigma_r = 2$, $k_r = 1$. For (D), (E), and (F), the initial condition is $(m, v) = (0, 0)$.

Asymmetric mood sensitivity to positive and negative events can lead to unipolar depression/mania

Asymmetric response to positive and negative events and its effects on human learning have been widely reported and inferred from psychological experiments (Leppänen, 2006; Pulcu & Browning, 2017). To investigate how this asymmetry can induce bipolar disorder within our model, we first compare the asymmetric learning case with experimental observations, especially in the case of unipolar depression and mania, characterized by a persistent deviation of mood in one direction from the baseline level. It has been observed that patients with major depression respond more strongly to negative stimuli than to positive stimuli (Gotlib, Kasch, et al., 2004; Gotlib, Krasnoperova, et al., 2004). Patients with mania, on
the other hand, show less response to negative stimuli (Lennox, Jacob, Calder, Lupson, & Bullmore, 2004). Interesting, patients with bipolar disorder, even in euthymic or depressive episodes, show stronger response to both positive and negative stimuli (Lawrence et al., 2004), consistent with our results in the previous subsection where the response is characterized by the mood sensitivity $f_{\eta m}$.

![Figure 3](image_url)

**Figure 3.** Response to jumps in reality with $r(t) = -4$ for $t \in [0, 5)$ and $r(t) = 4$ for $t \in [5, 10]$. Initial conditions are set to $(m, v) = (0, 0)$. Asymmetric mood sensitivity leads to unipolar depression/mania in our model. (A) The expectation $v$ in the normal state (solid blue), the manic state (red dash-dot), and the depressive state (green dotted) all attempt to follow reality (black dotted). In the depressive state, $v(t)$ overshoots decreases in $r(t)$, while expectations in the manic state overshoot rises in $r(t)$. The normal, manic, and depressive states are described by $f_{\eta m}^+/\eta_v + k = 0.4, 0.8, 0.1,$ and $f_{\eta m}^-/(\eta_v + k) = 0.4, 0.1, 0.8$, respectively. Numerical values for other parameters are $\eta_v = 0.37, f = 0.3, k = 0.37,$ and $k_3 = 2.8 \times 10^{-3}$. (B) Mood levels $m(t)$ exhibit significant systematic differences in the normal, manic, and depressive states, showing how asymmetric mood sensitivity can lead to unipolar depression/mania under changing reality. (C) Prolonged periods of negative mood are reflected by longer periods of large QIDS score in depressed individuals.

To see if our model can support unipolar depression/mania with experimentally observed asymmetry, we allow the learning rate in mood $\eta_m^\pm$ to have different values when experiencing positive and negative events (see Eq. (4) for details). In the $v - m$ plane this corresponds to different values of mood sensitivity in the different half-planes separated by the nullcline of $\frac{dv}{dt} = 0$, leading to a continuous but non-differentiable vector field. While the linear stability is not well-defined in certain parameter ranges (since the Jacobian is different on the two sides of the origin), it is clear that if both $f_{\eta m}^+ < \eta_v + k$ and $f_{\eta m}^- < \eta_v + k$ the origin is linearly stable (the non-bipolar state), and that if both $f_{\eta m}^+ > \eta_v + k$ and $f_{\eta m}^- > \eta_v + k$ the origin is linearly unstable leading to oscillations and some form of bipolar disorder.

The mathematical details of the general effects of asymmetry $\eta_{m}^+ \not= \eta_{m}^-$ have not yet been precisely determined.
Even under conditions of linear stability, the system may exhibit depression/mania under random reality because of the asymmetric response to changes in $r(t)$. A rigorous analysis of the systematic deviation of mood or expectation would require more involved stochastic analysis. However, we can easily simulate the dynamics of mood and expectation to strengthen our physical intuition about the effects of asymmetric mood sensitivity.

In Fig. 3 we show a simulation started well in the past with $r(t < 0) = 0$. The reality is then decreased to $r(0 \leq t < 5) = -4$, followed by an increase to $r(t \geq 5) = +4$. We see that the expectation of depressive individuals overreacts to negative reality and fails to fully recover by $t = 10$ when the reality switches positive at $t = 5$ (Fig. 3A). This lag in recovery leads to a prolonged time of depression compared to normal and manic individuals, reflected in both mood and QIDS score (Fig. 3B, C). The deviation in mood observed in the model can be explained in terms of psychology. Individuals with higher mood sensitivity for negative events will experience a larger change in mood during negative events, resulting in a lower expectation than reality. Since reality is typically changing, this overshoot in mood and expectation can last until the next event. Since individuals in depressive states will always overshoot in response to negative events and undershoot in response to positive ones, their overall mood level remains lower than those of normal individuals. A similar reasoning applies to people in a manic state, which results in average mood values higher than in normal subjects. Interestingly, our mechanism for unipolar depression/mania is distinct from another model based on the interaction of mood and expectation (Eldar et al., 2016), which asserts that asymmetric learning rates lead to expectations higher than reality for people with depression, resulting in constant negative surprise and low mood level. Our simulations show that the incorporation of experimentally observed asymmetric mood sensitivity can trigger unipolar depression/mania.

**Extreme asymmetry in mood sensitivity can lead to bipolar disorder**

Mathematically, bipolar disorder reveals itself in the form of a limit cycle as the origin $(m, v) = (0, 0)$ becomes linearly unstable. When asymmetric mood sensitivity is considered it is not straightforward to define the stability of the origin. However, it is easy to show that the origin remains stable if both positive and negative mood sensitivities are below the critical value, i.e. $f\eta^+_m, f\eta^-_m < \eta_v + k$. Similarly, the origin is unstable if both mood sensitivities are above the critical value. However, when only one of them is
Bipolar disorder can be triggered by extremely asymmetric mood sensitivity, even when one of the mood sensitivities does not cross the threshold. We observed that when the sum of the two mood sensitivities is lower than twice the critical value, a subject will remain in a normal state. When this sum is higher, a bifurcation to a bipolar state occurs. (A) Heat map of the standard deviation of the mood over the period \( t \in [81.25, 162.5] \). The simulations were performed within the period \( t \in [0, 162.5] \) using \( r(t > 0) = 0 \). Other parameter values used in the simulations are \( \eta_v = 0.37, f = 0.3, k = 0.37, \) and \( k_3 = 2.8 \times 10^{-3} \). The white solid line delineates the boundary between normal and bipolar states, characterized by the value when both mood sensitivities are critical: 
\[
 f(\eta_v + \eta_m) = 2 \times (0.37 + 0.37) = 1.48.
\]
(B) Under constant reality, bipolar disorder triggered by mood sensitivity asymmetry in different directions induce different behavior in expectation \( v(t) \). Compared to the normal state (blue solid), higher negative mood sensitivity (depressive bipolar state, \( f\eta_v = 2(\eta_v + k) \) and \( f\eta_m = 0.5(\eta_v + k) \)) lowers expectations (green dotted) while higher positive mood sensitivity (manic bipolar state, \( f\eta_v = 0.5(\eta_v + k) \) and \( f\eta_m = 2(\eta_v + k) \)) leads to higher expectations (red dash-dot). Initial conditions are \((m, v) = (0, -1)\).

(C) Under constant reality, bipolar disorder induced by asymmetry in mood sensitivities in different directions biases the mood \( m(t) \) in different directions. (D) The biases in the asymmetry-induced oscillations in the expectation persist under random reality conditions, with depressive/manic bipolar states leading to statistically lower/higher expectations. The realization of reality is drawn as described with \( \sigma_r = 2, k_r = 1 \). Initial conditions: \((m, v) = (0, 0)\).

(F) Predictions of QIDS scores of depressive and manic bipolar individuals.
that when the sum of positive and negative mood sensitivities is lower than twice the critical value $\eta_v + k$,
i.e. $f\eta_{m}^+ + f\eta_{m}^- < 2(\eta_v + k)$, bipolar disorder will not arise; otherwise, the mood oscillates and bipolar
disorder occurs. Consistent with our intuition, it is now the sum of both mood sensitivities that
determines the stability. While the intuition is clear, analytically calculating the properties of a limit cycle
in the presence of a stable half-plane is mathematically more involved and will be a topic of future
mathematical investigation.

Bipolar disorders triggered by asymmetric mood sensitivities show oscillation in mood and expectation
as in the symmetric case, but have systematic biases (Fig. 4B, C) which were not observed in the
symmetric case. As in unipolar depression/mania, the biases in mood and expectations always have the
same sign, i.e. either both mood and expectations are systematically lower or both higher. The
depression-biased case may describe type II bipolar disorder. The same pattern persists when the reality
is treated as random (Fig. 4D, E), with the mood and expectation responding to changes in reality as well
as exhibiting their intrinsic oscillations. As expected, the QIDS score for depressive bipolar (or type II)
individuals are much higher than those for normal and manic bipolar individuals, but even bipolar manic
individuals exhibit larger QIDS score than normal individuals (Fig. 4F). Moreover, manic and depressive
bipolar individuals often show high QIDS scores when normal individuals have stable mood. Our
numerics suggest that bipolar disorder can be caused by extreme asymmetry in mood sensitivity, which
leads to systematically biased mood and expectation patterns. The direction and magnitude of mood
sensitivity asymmetry may be an underlying feature of different types of asymmetric bipolar disorders.

**Effects of antidepressants and lithium**

In this section we explore the effects of common medications used to treat bipolar disorder. First we want
to see if our model can explain the antidepressant-induced mania seen in bipolar patients. Antidepressants
are a category of medicine for treating depression disorder, and their effects on patients with depression
are significant (Morris & Beck, 1974). For patients with bipolar disorder, it has been reported that
20–40% of the manic episodes are induced by antidepressants (Altshuler et al., 1995; Goldberg &
Truman, 2003). This unanticipated effect was previously studied by Goldbeter (2011) using a bistable
model of depression and mania. Our model for bipolar disorder is intrinsically oscillatory, and it is not
clear whether there is a threshold of dosage above which the manic episodes will be induced, as predicted
Figure 5. The effects of antidepressants and lithium on patients with bipolar disorder, including the mania-inducing effect of antidepressants and the sedative effects of lithium, are assessed in our model. (A) Numerical simulation of the mood of a bipolar patient (solid blue curve) using $f_{\eta^m} = 1.5(\eta_v + k)$. At $t = 9.2$ weeks, within a depressive episode, the patient is treated with antidepressants, modeled by an elevation in mood (Goldbeter, 2011). Trajectories corresponding to dosages that instantaneously decrease the depression to 70% of its lowest value (green dotted), 30% of its lowest value (red dash-dot), and 10% of its lowest value (black dotted) are shown. Note that higher doses lead to an earlier onset of mania. This antidepressant-induced mania is observed clinically (Altshuler et al., 1995; Goldberg & Truman, 2003). The numerical values for the simulations are $\eta_v = 0.37$, $f = 0.3$, $k = 0.37$, and $k_3 = 2.8 \times 10^{-3}$ and the initial conditions are $(m, v) = (0, -1)$. (B) The quick transition to a manic phase results in a depressive episode that occurs sooner than untreated individuals, as indicated by an earlier peak in QIDS score for individuals treated with a high antidepressant dose. (C) Simulated mood dynamics for mania-biased mood sensitivity asymmetry (red dotted, $f_{\eta^m} = 1.5(\eta_v + k)$, $f_{\eta^-} = (\eta_v + k)$) and depression-biased mood sensitivity asymmetry (blue solid, $f_{\eta^m} = (\eta_v + k)$, $f_{\eta^-} = 1.5(\eta_v + k)$). The sedative effects of lithium are modeled via a symmetric 20% reduction in mood sensitivity and is implemented in our simulations at $t = 27.1$ weeks (black arrow). This treatment decreases oscillation amplitudes consistent with clinical observations (Phiel & Klein, 2001). (D) The reduction in mood oscillation amplitudes yields smaller predicted QIDS scores.

in Goldbeter (2011). Nevertheless, simulations of our model show that there is, indeed, a threshold of dosage under which a transient alleviation of depression occurs followed by a usual manic episode. Above this threshold of medication, manic episodes are induced earlier (Fig. 5A). This result is surprising since small perturbations in mood do not qualitatively change the subsequent dynamics and our model does not have a built-in mechanism for bistability. Individuals treated with high doses of antidepressants are predicted to show a phase shift in the mood oscillations (Fig. 5A). This phase shift would yield an earlier peak in the QIDS score (Fig. 5B). While certain clinical observations such as rapid cycling are not
observed (Altshuler et al., 1995), our model reproduces antidepressant-induced manic episodes and exhibits behavior that is consistent with previous models relying on bistability (Goldbeter, 2011).

The sedative effects of lithium was first discovered in 1949 although its molecular mechanisms of action have not yet been fully elucidated (Corbella & Vieta, 2003; Phiel & Klein, 2001). Notwithstanding, lithium is one of the most prescribed treatments for bipolar disorder (Phiel & Klein, 2001). While our model does not explicitly involve details at the molecular level, it suggests a crucial behavioral property, characterized by the mood sensitivity parameter, that might be regulated by lithium. To see this, we simulate the mood \( m(t) \) in the bipolar state and decrease the mood sensitivity parameter after a certain time point (Fig. 5C). We observe that after the mood sensitivity is decreased, the amplitudes of oscillations in mood gradually decrease, eventually becoming constant over time. Depression is lessened after treatment, as indicated by a decrease in the QIDS score (Fig. 5D); moreover, decreases in mood sensitivity do not induce mania. In contrast to antidepressants, lithium does not trigger manic episodes, which makes it suitable to treat bipolar depression (Phiel & Klein, 2001). This result suggests that the sedative effect of lithium might be achieved by decreasing the mood sensitivity parameter rather than directly modifying mood.

**DISCUSSION AND CONCLUSION**

In this work we propose and analyze a simple mathematical model for bipolar disorder based on the interaction between mood and expectation. Our model exhibits oscillatory mood behavior when the mood sensitivity parameter exceeds a threshold. Previous models have explained such oscillations via the dynamics of intrinsic brain circuits or mutual inhibition of depression and mania (Bonsall et al., 2015; Daugherty et al., 2009; Goldbeter, 2011). Our model proposes that mood oscillations arise from a psychological mechanism in which high expectation induces high mood until it reaches a physiological limit. The mood then decreases, followed by a concomitant decrease in expectation. This mechanism is similar to that proposed by Eldar et al. (2016) and Mason et al. (2017), but in our model, we identify a key psychological property, defined by the mood sensitivity parameters \( \eta_{m}^{\pm} \), that may control a whole spectrum of states, from normal, to cyclothymic personality, to type I and type II bipolar disorders. Measuring mood sensitivity may result in a more general diagnostic method to classify and describe such disorders.
We also explored the effect of asymmetric mood sensitivity on unipolar depression/mania and bipolar disorder. Humans are known to react differently toward positive and negative events (Pulcu & Browning, 2017) and patients with major depression and bipolar disorder have stronger bias toward these events (Leppänen, 2006). It has been suggested by Eldar et al. (2016) that this asymmetry can lead to unrealistic expectation and low mood in depressive patients. Our analysis shows that depression can result from a higher mood sensitivity toward negative events, which leads to a reasonable expectation but negative mood. Our model also predicts that depression is a dynamical phenomenon, i.e. when no strong environmental stimulus is present, depressive patients may appear normal but they react more negatively than normal subjects once reality fluctuates. Our prediction is supported by clinically observed processing bias (Fu et al., 2004) but additional psychological experiments should be performed to test our model hypotheses.

Our model is among the first to show that extreme asymmetry in mood sensitivity can lead to full-on bipolar disorder. Here, the mood sensitivity plays the role of a force that pushes mood and expectation away from their balanced-state values. Bipolar disorder will occur as long as the sum of forces toward positive and negative stimuli is larger than a certain threshold. Our mathematical framework can explain a paradoxical observation that while depressive patients react more strongly to negative events, bipolar patients in the depressive phase can show stronger reactions to positive events (Lawrence et al., 2004; Leppänen, 2006).

Asymmetry in the mood sensitivity parameter introduces an interesting mathematical question on stability. Conventionally, the local stability of an equilibrium is determined by the stability of the system linearized around the equilibrium (Strogatz, 2014), with certain exceptions in which the full nonlinear system has to be considered. We showed numerically that if the linear stability differs on two half-planes splitting the equilibrium, then the equilibrium can be either stable or unstable. It is intuitively clear that the stability depends on the average effect of both half-planes, but there has been no rigorous treatment of this thus far.

To the best of our knowledge, existing models for bipolar disorder are based on one of two basic mechanisms: biological rhythm and bistability. The first mechanism models the oscillation in mood as a consequence of an intrinsic oscillation in the brain. In this case, mood oscillations persist without perturbations (Bonsall et al., 2015; Daugherty et al., 2009; Goldbeter, 2011; Mason et al., 2017). The
second mechanism assumes that there are multiple stable states representing a spectrum of depression
and mania. Here, variations in mood are triggered by random external perturbations arising from life
events (Cochran et al., 2017; Steinacher & Wright, 2013). The perturbations from life events in biological
rhythm models are usually treated as a noise term in the oscillator. We have modeled life events explicitly
by a time-dependent reality function \( r(t) \) which enables a direct comparison of the two mechanisms since
different forms of \( r(t) \) can be used to investigate which mechanism better explains the observations. For
example, when an individual experiences a prolonged negative life event, biological rhythm models
would predict a persistence in mood oscillation, while bistability models would likely predict a prolonged
state of depression. By directly incorporating reality \( r(t) \) into models with different central mechanisms
and then comparing their predictions with observations, we may be able to decide which model better
describes bipolar disorder. This may also reveal a need for combination of different mechanisms.

Due to a lack of understanding of the underlying biological mechanism of bipolar disorder, the
parameters in models for bipolar disorder are often phenomenological and treated as fitting parameters to
the experimental data. Our model cannot avoid phenomenological parameters in the sense that they
cannot be measured from biological experiments. However, our parameters are expressed in
psychological terms, such as learning rate for expectation, recovery rate for mood, etc., that can be
measured by psychological experiments instead of fitting to data. For example, reaction toward events
can be measured by fMRI or pupilometry (Fu et al., 2004; Lawrence et al., 2004; Pulcu & Browning,
2017) which can then be used to estimate the learning rates and the mood sensitivity parameter. In fact,
the measurements of Lawrence et al. (2004) showing that bipolar patients react more strongly to both
positive and negative events agree with the prediction of our model. The ability to define and measure our
parameters from psychological experiments will render our model and predictions falsifiable. We expect
our model to provide a basis for quantitative diagnosis of bipolar disorder and to guide the development
of new treatments and therapies.

In its current form, our model is defined by parameters that are constant in time. In reality, these
parameters might interact with mood and expectation, making them part of the dynamical system. At the
cell level, neural synapses can be modified by the synaptic current (Fain, 1999), which suggests that
recurrence of negative events might strengthen reactions to them. It has been observed that depression is
correlated to chronicle pain (Geisser, Roth, Theisen, Robinson, & Riley III, 2000) and that an initial
depression might become long-term because of environmental difficulties and lack of social support, the so-called cognitive vulnerability (Persons & Miranda, 1992). This evidence suggests that the psychological parameters in our model are dynamical and affected by the environment instead of heredity. Therefore, a natural next step in our work is to incorporate the dynamics of mood sensitivity as well as other parameters to see if the recurrences of external events can trigger depression/mania or bipolar disorder. The design and analysis of a model with dynamical parameters is a rich topic for future study.

ACKNOWLEDGMENTS

The authors are grateful for support from the Army Research Office (W911NF-14-1-0472) and the National Science Foundation (DMS-1516675). SSC was supported in part by the Systems and Integrative Biology predcotoral training grant (T32GM008185).

REFERENCES


