



OPEN

## Analyzing the effects of memory biases and mood disorders on social performance

Nanda Kishore Sreenivas<sup>1,3</sup> & Shrisha Rao<sup>2,3</sup>✉

Realistic models of decision-making and social interactions, considering the nature of memory and biases, continue to be an area of immense interest. Emotion and mood are a couple of key factors that play a major role in decisions, nature of social interactions, size of the social network, and the level of engagement. Most of the prior work in this direction focused on a single trait, behavior, or bias. However, this work builds an integrated model that considers multiple traits such as loneliness, the drive to interact, the memory, and mood biases in an agent. The agent system comprises of rational, manic, depressed, and bipolar agents. The system is modeled with an interconnected network, and the size of the personal network of each agent is based on its nature. We consider a game of iterated interactions where an agent cooperates based on its past experiences with the other agent. Through simulation, the effects of various biases and comparative performances of agent types is analyzed. Taking the performance of rational agents as the baseline, manic agents do much better, and depressed agents do much worse. The payoffs also exhibit an almost-linear relationship with the extent of mania. It is also observed that agents with stronger memory perform better than those with weaker memory. For rational agents, there is no significant difference between agents with a positive bias and those with a negative bias. Positive bias is linked with higher payoffs in manic and bipolar agents. In depressed agents, negative bias is linked with higher payoffs. In manic agents, an intermediate value of mood dependence offers the highest payoff. But the opposite is seen in depressed agents. In bipolar agents, agents with weak mood dependence perform the best.

Individual and group decision-making and social interactions have been studied for long and continue to remain of interest. Some early models were fairly simplistic<sup>1,2</sup>, and they assumed perfectly rational behavior which is often not the case in reality. Time and again, research in the domains of psychology and behavioral sciences asserts that humans, as well as animals, are subject to a wide array of cognitive biases<sup>3,4</sup>. Later works moved away from the assumption of perfect rationality and started incorporating various cognitive biases into decision models<sup>5-8</sup>. However, the effects of emotion and mood disorders on decisions and interactions still need to be explored.

Although we humans aspire to be rational decision-makers and never let emotion cloud our judgment, such is never really the case<sup>9</sup>. This has been understood for long—the importance of emotion was captured succinctly by Simon<sup>10</sup>:

“Hence, in order to have anything like a complete theory of human rationality, we have to understand what role emotion plays in it.”

A great deal of research has focused on the impact of emotion on memory but the findings have been extremely diverse, with some claiming emotional memories are indelible<sup>11</sup>, while others claim that emotion has no effect whatsoever on memory<sup>12</sup>. Even among those that agree that emotion has some impact on memory, some claim memories associated with positive emotions are stronger<sup>13,14</sup> and last longer, while others claim the opposite<sup>15</sup>.

Appraisal theory has been used in prior works to quantitatively define emotion, where an agent evaluates the reward with respect to its goal or expectation<sup>16,17</sup>. If the reward obtained is better than the agent's expectation, positive emotion is induced; and the reverse also holds. The process of storing information for the first time in memory is called *encoding* and this process leaves a *trace* in the memory. The strength and longevity of the trace depend on how well it has been encoded.

Mood has been seen as the summary of recent emotions<sup>18</sup>. While emotions change after each episode, mood maintains a historical context, yet staying temporally relevant. Mood affects the behavior of an individual in terms

<sup>1</sup>Oracle, Bangalore, India. <sup>2</sup>International Institute of Information Technology - Bangalore, Bangalore, India. <sup>3</sup>These authors contributed equally: Nanda Kishore Sreenivas and Shrisha Rao. ✉email: shrao@ieee.org

of reward perception<sup>19</sup>, memory retrieval<sup>20</sup>, and levels of social engagement<sup>21,22</sup>. There is much to be learned about the effects of emotions and mood in decisions and social interactions. Shevchenko<sup>23</sup> recently wrote:

“Moods in decision research have been ignored for many years, although decisions made in an emotional state are ubiquitous in daily life. The areas of emotion and decision-making research have remained mostly unfamiliar.”

Mood disorders are prevalent in humans, and studies indicate that about 20% of U.S adults experience some class of mood disorder in their lifetimes<sup>24</sup>. There have been many computational models of various mood disorders<sup>19,25–28</sup>, but most of these assume independent agents without any social interactions or social networks. The extent and frequency of social interactions are intricately linked to mood disorders, with studies indicating that chronically depressed people have smaller social networks than healthy controls<sup>29,30</sup>. Empirical studies also conclude that depressed phases are connected with lower energy, interest, and drive<sup>31,32</sup>. Prior models do not take this into account.

There are several models of mood fluctuation in bipolar disorder such as rhythmic models, discrete models, and models based on the Behavioral Approach System (BAS)<sup>33–37</sup>.

Symptoms of depression include persistent low mood or sadness, and an inability or diminished ability to experience pleasurable stimuli. Depression and mania have been modeled using the concept of *reward sensitivity* in some prior works<sup>19,26,27</sup>. In these models, the reward as perceived by an agent is the product of its reward sensitivity and the actual reward. Depressed agents are assigned values of reward sensitivity between 0 and 1, while manic agents have values greater than 1.

Mood also affects memory retrieval; *mood dependence* is the phenomenon where past events whose emotional state match the current mood are more likely to be retrieved. There seems to be no prior model capturing this effect of mood on memory.

In some works on decision-making, memory is assumed to be infinitely long and perfect<sup>5,38</sup>. However, this is not justified<sup>39,40</sup>, and memories are susceptible to forgetting at various rates. The likelihood of remembering recent events is higher when compared with older events, which is termed the “recency effect”<sup>41,42</sup>. While it seems intuitive that stronger memory enhances performance, some have also argued that forgetting is a useful feature because it helps optimize decision-making<sup>43</sup>. The most common approaches to model forgetting involve the use of exponential decay or power law functions<sup>44–46</sup>.

This work is broadly in the field of computational social science using agent-based modeling<sup>47</sup>, and specifically uses an agent-based model carefully constructed in line with studies in psychology. It is well known that agent-based modeling is a powerful, relatively new tool in the study of social psychology<sup>48</sup>, and that while we cannot yet directly derive outcomes that can guide treatments, it does permit new insights that are not possible using mathematical models and clinical experiments<sup>49</sup>.

We give an integrated model of decisions and social interactions that takes into account the roles of emotion, mood, and memory biases. A game of iterated interactions based on the Continuous Prisoners’ Dilemma (CPD) is considered. As opposed to a standard Iterated Prisoners’ Dilemma which offers only two choices, CPD allows agents to cooperate at various levels between 0 and 1. To explain one such interaction, assume an agent *A* is paired with one of its neighbors *B*. *A* appraises the reward, i.e., the cooperation extended by *B*, with respect to its expectation, and this determines the emotion induced by this interaction. Both agents update their memory and mood, and this affects their subsequent interactions. Agents also broadcast the gist of their interaction to their neighbors.

In our work, emotion is modeled in line with the appraisal theory. To model the effect of emotion on memory, we define the initial strength of a trace as a linear function of emotion. We use two different gradients for positive and negative emotions, allowing us to model both positively and negatively biased agents.

The mood dependence effect is modeled as a probabilistic retrieval of traces using a triangular distribution centered around the current mood of the agent. This ensures traces with similar emotional states have higher chances of being retrieved than others, which is the condition of mood dependence. Forgetting is modeled using exponential decay, in line with prior work<sup>45,46</sup>.

Agents with mood disorders such as mania, depression, and bipolarity are considered in addition to perfectly rational agents. Such an agent society is modeled with an interconnected social network, where agents interact only with their neighbors, and agents broadcast their experiences to their neighbors. Also, the size of the personal network of each agent is determined by its type, which helps fill one significant gap in many prior models that consider agents in isolation.

Agents are also modeled with different levels of engagement or drive. In each round of interaction, an agent has the choice to decide if it wants to participate. Manic agents are modeled with a higher drive and depressed agents with a lower drive, as is consistent with clinical observations<sup>50,51</sup>.

Through simulation, we obtain results about the comparative performances of different types. An agent society with agents of all four types—rational, manic, depressed, and bipolar—is considered. Considering the average payoff of rational agents as the baseline, bipolar agents perform better by roughly 10%. The average payoff of manic agents is approximately 70% higher, and it is about 60% lower for depressed agents. Average payoff is plotted against the extent of mania and depression, and an almost-linear relationship is observed in both cases.

With regard to memory, agents with stronger memory perform better than those with weaker memory. The results differ across types with regard to emotional bias. For rational agents, there is no significant difference between agents with a positive bias and those with a negative bias. In manic agents as well as bipolar agents, positive bias is linked with higher payoff, but the effect is larger for bipolar agents when compared to manic agents. However, among depressed agents, those with a negative bias have higher payoffs.

In the context of mood-dependent retrieval also, the effects vary across different types. For rational agents, there is no significant change in payoff observed by varying the level of mood dependence. In manic agents, an intermediate value of mood dependence offers the highest payoff. But in depressed agents, a similar medium value offers the least payoff. In bipolar agents, agents with weak mood dependence perform the best.

## Background

In this section, we summarize related work in the fields of psychology, decision and behavioral sciences. Each subsection deals with one key aspect of the context of our model.

**Forgetting.** Forgetting is the time decay of strength of memory traces in the absence of refreshing, until it is ultimately deleted. This is a fairly common phenomenon in humans, and has a considerable impact in decision-making. The earliest work on memory and forgetting can be traced back to 1885, when Ebbinghaus examined the decline of memory strength with time. Broadly speaking, there have been two theories to model forgetting, on the basis of decay and interference<sup>52–54</sup>.

While both theories agree that old memories are lost over time, decay theory posits time as the only driving force behind forgetting<sup>55,56</sup>. Recent models based on time decay include the primacy model<sup>57</sup> and the positional decay model<sup>58</sup>. Interference-based models argue that other activities interfere with retrieval of memory items, and thus contribute to forgetting. Typically, these models use an associative network that binds items to contexts, and when new information is encoded it typically overwrites the old, which is thus forgotten<sup>59</sup>.

Rubin and Wenzel considered two hundred datasets and about a hundred different functions, and observed that logarithmic, exponential, hyperbolic and power functions offered the best fit to model forgetting<sup>60</sup>. Since then, exponential decay<sup>44–46</sup> and power functions<sup>61–63</sup> seem to be the most favored functions to model forgetting.

**Emotions and memory.** Levine and Pizarro write<sup>64</sup>: “Converging evidence from autobiographical memory studies, animal and human laboratory studies, and brain imaging studies shows that emotional events are remembered better than non-emotional events and that mechanisms specific to emotion underlie these effects.” Several other studies also concur with the idea that emotions improve memory for details central to the event and diminish memory for any peripheral details<sup>65,66</sup>.

In the domain of modeling emotions, one of the most influential ideas is the appraisal theory in psychology, where emotions are based on an individual’s evaluation of an interaction with respect to a goal<sup>16,17</sup>. There have been attempts to relate the emotional valence and strength of autobiographical memories, but the findings are varied, with some claiming that positive events are remembered better<sup>13,14</sup>, while others support better memory for negative events<sup>15</sup>.

Negativity bias has been described as the phenomenon by which a negative event has a greater psychological effect even when it is of equal intensity as a positive event<sup>67</sup>. This bias has been found to be fairly ubiquitous in humans<sup>68,69</sup>. Neurological studies also point to greater processing of negative information, which causes asymmetry in attention and memory<sup>70</sup>. In studies concerning positive and negative behaviors, participants tend to recall negative behaviors more than positive behaviors<sup>71,72</sup>.

**Mood and memory.** It has been known for long that mood and memory are related, and that the current mood affects which memories are retrieved<sup>73,74</sup>. As Lewis et al.<sup>20</sup> write, “Remembering all of the negative events of our past lives when depressed is an example of this.” Mood dependence is the phenomenon where retrieval is facilitated for memories that match the present mood. Neuroimaging studies examining the influence of emotion in encoding and retrieval also provide evidence in support of this<sup>75,76</sup>. Further, recent empirical studies based on human participants also offer support to this theory<sup>77</sup>. However, there seem to be no computational models of this phenomenon.

**Types of mood disorders.** *Bipolar disorder.* Bipolar disorder<sup>78</sup> (BP) is a mental disorder characterized by cycling between manic and depressed episodes. Its prevalence is estimated to be 3.3% in the US<sup>79</sup>, and it is thought to affect about 2.4% of the global adult population<sup>80</sup>.

There are several prior computational models of BP. Rhythmic models are based on the idea that there is an intrinsic oscillation of mood, independent of any external perturbation<sup>33,34,81</sup>. These models assume mood swings periodically between mania and depression<sup>82,83</sup>. Another class of models assume multistability, claiming mood tends to distinct discrete states<sup>35,84</sup>. The other popular theory is that the hypersensitivity of the Behavioral Approach System (BAS), and the resulting interactions between mood, expectation, and behavior can explain BP<sup>36,37,85</sup>.

However, among these classes of models, rhythmic models are thought to have close links to actual biological rhythms, such as the circadian rhythm<sup>86</sup>. Indeed, the standard clinical treatment for BP, using Lithium, works through influencing the circadian pathways<sup>87</sup>.

*Depression and mania.* Anhedonia is one of the major symptoms of depression, and it refers to an inability or a diminished ability to experience pleasurable stimuli<sup>51,88</sup>. This model is supported by an empirical study as well, in which self-reported anhedonia was positively correlated with participants’ estimated reward sensitivity<sup>89</sup>. Mania is characterised by elevated mood, energy, and reward-seeking behavior or drive<sup>90</sup>.

A popular approach to model mania and depression is to use reward sensitivity. The perceived reward is determined by the product of actual reward and reward sensitivity<sup>19</sup>; to model mania, reward sensitivity is chosen

as a value greater than 1, and to model depression, a fractional value is used. Thus, the single parameter can be used to model diminished and enhanced reward perception in depression and mania respectively.

Social interactions in mania can be abnormal, such as acting in overly familiar ways with strangers<sup>50</sup>. Depressed agents on the other hand exhibit poor social relationships, with studies indicating lower scores for adequacy of both intimate and diffuse social relationships<sup>32,91,92</sup>. Studies also indicate elevated absenteeism at work, and lower work performance, in chronically-depressed individuals, and during depressed phases in BP patients<sup>31</sup>.

### Model of agent psychology and interactions

Consider two agents *A* and *B* paired in an interaction at time *t*. Now, agent *A*'s expectation of *B*'s level of cooperation in this interaction ( $E_A(B)$ ) is the average of past levels of cooperation of *B*. *A* searches its memory for past experiences with *B*, which returns a set of relevant traces denoted by  $L_e$ .

$$E_A(B) = \frac{\sum_{i \in L_e} i[c]}{|L_e|}$$

Here,  $i[c]$  denotes the level of cooperation extended by the other player in the interaction that is described by the trace *i*. The primary assumption of this model, in line with studies<sup>93,94</sup>, is that *A* cooperates at the same level at which it expects *B* will cooperate.

$$c_A = E_A(B). \tag{1}$$

This interaction induces an emotion in both agents, and consequently their moods and memory are updated as well. These will be explained in more depth in the following subsections. For the remainder of this section, we will continue with the same nomenclature of agents, *A* and *B* to maintain consistency and clarity.

**Interaction model.** The agent society consists of four different types of agents: rational, manic, depressed, and bipolar agents. The set of all agents is denoted by  $\mathcal{A}$ . The society is modeled with a social network, and agents' interactions are limited to their neighbors only. The social network is modeled as an undirected graph with agents as the nodes. Hence, the social network can be represented by

$$G = (\mathcal{A}, \mathcal{E}) \quad \mathcal{E} \subseteq \{(x, y) | (x, y) \in \mathcal{A}^2 \wedge x \neq y\}.$$

The degree of the node determines the size of the neighborhood of an agent. The size of the neighborhood of the agent *A* is given by  $\eta_A$ , so  $0 \leq \eta_A \leq |\mathcal{A}| - 1$ .

As discussed in "Types of mood disorders", the size of the neighborhood of an agent is dependent on the type of the agent. The exact relationship between these is explained in "Modeling mood disorders".

The simulation consists of multiple rounds of interactions, and all agents are not active in all the rounds. At any such round *t*, an agent *A* is randomly paired with another agent *B* if and only if  $(A, B) \in \mathcal{E}$  and both *A* and *B* are active at *t*.

An interaction between a pair of agents is based on the Continuous Prisoners Dilemma (CPD)<sup>95,96</sup>. In the standard Iterated Prisoners' Dilemma, agents are restricted to only two actions—cooperate or defect. However, not all interactions can be realistically modeled with such restricted behavior.

In the CPD, agents can cooperate at any level between 0 and 1, where 0 and 1 correspond to the cases of completed defection and cooperation respectively. The concept and related payoff structure are adopted from Verhoeff<sup>95</sup>.

For the two agents *A* and *B*, their cooperation levels are denoted by  $c_A$  and  $c_B$ , respectively. The payoff functions are derived from the discrete payoff matrix by linear interpolation<sup>95</sup>:

$$p_A(a, b) = c_A c_B C + c_A \bar{c}_B S + \bar{c}_A c_B T + \bar{c}_A \bar{c}_B D, \tag{2}$$

where *C*, *T*, *D*, *S* are the payoffs in the standard PD as shown below.

		Player <i>B</i>	
		1	0
Player <i>A</i>	1	( <i>C</i> , <i>C</i> )	( <i>S</i> , <i>T</i> )
	0	( <i>T</i> , <i>S</i> )	( <i>D</i> , <i>D</i> )

The values are chosen so  $2C > T + D$  and  $T > C > D > S$  (these are standard conditions applicable to the Prisoners' Dilemma).

**Emotion and mood.** We adopt the appraisal theory of emotion, as discussed in "Emotions and memory". The agent *A* evaluates the reward, which is the level of cooperation  $c_B$  extended by *B* with respect to its expectation  $E_A(B)$ . The emotion of the agent *A* at time *t*,  $\delta_A^t$  is therefore the difference between reward and expectation, given by

$$\delta_A^t = c_B - E_A(B) = c_B - c_A. \tag{3}$$

By Eq. (1),  $E_A(B) = c_A$  and thus the induced emotion is equivalent to the difference between the levels of cooperation of  $A$  and  $B$ . As discussed in "Interaction model", levels of cooperation  $c_A$  and  $c_B$  range from 0 to 1. Therefore Eq. (3) shows that  $-1 \leq \delta_A^t \leq 1$ . If  $c_B > c_A$ , it implies  $A$  has received a reward greater than it expected, which results in a positive emotion,  $\delta_A^t > 0$ .

The mood  $v_A^t$  of agent  $A$  is updated as follows.

$$v_A^t = \tanh(v_A^{t-1} + \kappa_A(\delta_A^t - v_A^{t-1})). \quad (4)$$

Here,  $\kappa_A$  is the rate of change of mood and  $0 \leq \kappa_A \leq 1$ . When  $\kappa_A$  is 1, the agent's mood is fleeting and is the same as its emotion. But, when  $\kappa_A$  is 0, the agent's mood is constant and remains unperturbed by the interactions. To maintain symmetry, the mood is transformed to lie between  $-1$  and  $1$  using a sigmoidal function such as the hyperbolic tangent function.

**Memory encoding.** The agent stores its experience after each interaction in its memory. This process of encoding leaves behind a trace, with initial strength  $\phi_0$ . We model that the encoding function is solely based on the emotional value of the interaction. As discussed in "Emotions and memory" and the references therein, it is clear that any encoding function satisfies the following two properties. First, emotionally charged events must be more strongly encoded than emotionally neutral events. Second, any agent can either be positively or negatively biased and the encoding function should allow this. After the interaction between  $A$  and  $B$ ,  $A$  now encodes this experience into its memory, and the initial strength  $\phi_0$  of this trace is given by:

$$\phi_0 = \begin{cases} (-k_N \times \delta_A^t) + c & \delta_A^t \leq 0 \\ (k_P \times \delta_A^t) + c & \delta_A^t > 0 \end{cases} \quad (5)$$

The negative and positive gradients are denoted by  $k_N$  and  $k_P$ . It can be seen from Eq. (5) that  $c \leq \phi_0 \leq \max(k_N, k_P)$ . The values of the gradients may differ across agents, but for any agent it remains constant. The emotional bias of an agent, denoted by  $\xi$ , is a ratio of the two gradients.

$$\xi = \frac{k_P}{k_N}. \quad (6)$$

If  $\xi > 1$ , then the agent is positively biased; it is negatively biased if  $\xi < 1$ , and has no emotional bias otherwise.

**Rate of forgetting.** A trace cannot remain at the same strength as time progresses. We use an exponential forgetting curve as in many prior works<sup>44–46</sup>.

$$\phi_t = \phi_0 e^{-\theta_A t}. \quad (7)$$

Here,  $\theta_A$  is the rate of forgetting and varies across agents, but remains constant for an agent;  $\theta > 0$ . Each agent also has a minimum threshold on strength of traces,  $\phi_{\min}$ . When the strength of a trace drops below this threshold, it is deleted from memory.

However, when a trace is successfully retrieved from memory, its strength is reset to  $\phi_0$ .

**Remark 1** The duration  $d$  of availability of a trace, without any refreshing, increases with initial strength, and decreases with the rate of forgetting.

For a trace to get deleted from memory, its strength should be less than the minimum threshold, i.e.,  $\phi_d < \phi_{\min}$ . This means  $\phi_0 e^{-\theta d} < \phi_{\min}$ , which in turn yields:

$$d > \frac{1}{\theta} \ln \left( \frac{\phi_0}{\phi_{\min}} \right). \quad (8)$$

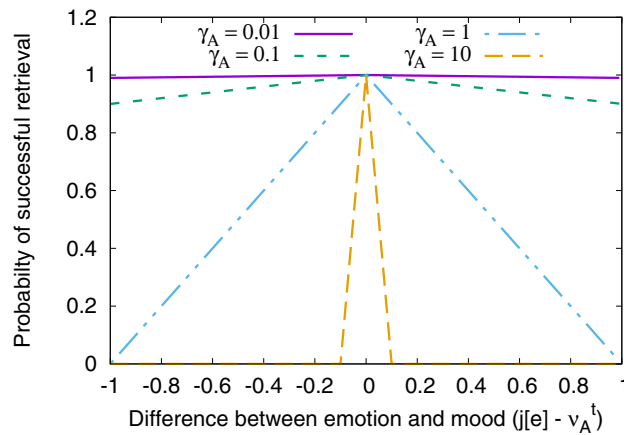
Since  $\phi_0$  is higher for emotionally charged events by Eqs. (5), (8) shows that emotionally charged events last longer in the memory.

**Mood and memory retrieval.** Agent  $A$  searches its memory for traces pertaining to  $B$ , and obtains the filtered set of traces,  $\mathbf{L}$ . Mood dependent retrieval is the phenomenon where traces whose emotional states match the current mood are more likely to be retrieved. Since mood  $v_A^t$  is a continuous variable such that  $-1 \leq v_A^t \leq 1$  by Eq. (4), retrieving traces whose emotional states are exactly equal to the current mood is not feasible. Therefore, we use a triangular function (similar to the triangular apodization function used in signal processing and spectroscopy<sup>97</sup>), centered at the agent's current mood, with a parameter  $\gamma_A \geq 0$ , to assign the likelihood of retrieval of each trace. Any function chosen should satisfy the following two requirements. First, for lower values of  $\gamma_A$ , the retrieval should not be selective, i.e., all traces should be retrieved irrespective of emotional state. Second, for higher values of  $\gamma_A$ , the retrieval chances should be better for those traces whose emotional state match the current mood. A triangular function satisfies both these requirements.

From the filtered set  $\mathbf{L}$ , the likelihood of retrieval of each trace is given by:

$$\forall j \in \mathbf{L}, p_j = \max(0, 1 - \gamma_A |j[e] - v_A^t|). \quad (9)$$

Here, emotion of trace  $j$  is denoted by  $j[e]$  and by Eq. (9), it is clear that  $0 \leq p_j \leq 1$ . When  $\gamma_A$  is lower, the curve is flatter, and almost all traces are equally likely to be retrieved with probability tending to 1, which is the



**Figure 1.** Mood dependence.

case of weak mood dependence. On the other hand, if  $A$  has a strong mood dependence, i.e., high  $\gamma_A$ , only traces whose emotional states are close to the current mood are retrieved. This is also seen in Fig. 1, where the curves corresponding to lower values of  $\gamma_A$  are almost horizontal and assign probabilities close to 1 for all traces. On the other hand, the curves corresponding to higher values of  $\gamma_A$  assign a probability of 0 to traces whose emotions do not match the current mood. Traces in  $L$  are probabilistically retrieved based on Eq. (9), and those traces that are successfully retrieved are returned as  $L_e$ .

**Modeling mood disorders.** We model three different mood disorders: mania, depression, and bipolar disorder.

*Mania and depression.* As discussed in "Types of mood disorders", one of depression's common symptoms is anhedonia<sup>51</sup>, the inability or the diminished ability to enjoy pleasurable events; in mania, the opposite is true<sup>90</sup>. Huys et al.<sup>19</sup> used reward sensitivity to model perceived rewards to be different from actual rewards in individuals with mood disorders. In Eq. (3), we assume no difference between perceived and actual rewards, both equal to  $c_B$ . To accommodate manic and depressed agents, Eq. (3) is modified as:

$$\delta_A^t = \tanh(\beta_A^t c_B - c_A). \tag{10}$$

Here,  $\beta_A^t$  is the reward sensitivity of agent  $A$  at time  $t$ , which remains constant for each agent, of either manic and depressed type. If agent  $A$  is depressed, then  $0 \leq \beta_A \leq 1$ . If  $A$  is manic, then  $\beta_A > 1$ . When  $\beta_A$  is exactly 1, Eq. (10) reduces to Eq. (3), and  $A$  is a rational agent. Although theoretically  $0 \leq \beta_A < \infty$ , for the sake of simulation a maximum value of  $\beta_A$  is considered, and denoted by  $\beta_{\max}$ .

Depression is associated with reward hyposensitivity and anhedonia<sup>89,98</sup>, while mania is associated with excessive motivation<sup>99</sup>. Therefore we model reward sensitivity as a constant, trait-like parameter (to address hyposensitivity) with a low value (for anhedonia) for depressed agents, and a high value (for the opposite, in case of manic agents).

The mood  $v_A^t$  is initialized with a value based on  $\beta_A^0$  as:

$$v_A^0 = \begin{cases} \beta_A^0 - 1 & 0 \leq \beta_A^0 < 1 \\ 1 + \frac{\beta_A^0 - 1}{\beta_{\max} - 1} & 1 < \beta_A^0 \leq \beta_{\max} \end{cases} \tag{11}$$

It can be seen from Eq. (11), depressed agents are initialized with a negative mood, and manic agents are associated with a positive mood.  $0 \leq v_A^t \leq 1$  for manic agents and  $-1 \leq v_A^t \leq 1$  for depressed agents.

In the case of depression, it has been seen that depressed people are more likely to be connected to other depressed individuals<sup>100</sup>. In line with this, we introduce a connection diversity parameter for depressed agents only, denoted by  $\pi$ . If  $\pi$  is 0.5, that implies half the connections of the agent are depressed agents, and the remaining connections are distributed among the other types equally.

*Bipolar disorder.* Bipolar disorder (BP) is a mood disorder where the agent's mood oscillates between mania and depression. In line with existing rhythmic models<sup>34,81,101</sup> that assume mood swings as periodic, we model mood fluctuation in BP using a sine curve with the period varied periodically by the parameter  $\omega_A$  and a constant offset  $\lambda_A$ . The mood of a bipolar agent  $A$  at any time  $t$  is given by:

$$v_A^t = \sin \omega_A t + \lambda_A. \tag{12}$$

The offset  $\lambda_A$  determines if the agent is mostly manic or mostly depressed and  $-1 \leq \lambda_A \leq 1$ . If  $\lambda_A > 0$ , the agent is mostly manic it is mostly depressed if  $\lambda_A < 0$ . The parameter  $\omega_A$  determines the period of cycling. Clinical

studies suggest there are phases of rapid cycling in BP patients<sup>82,83</sup>, and that the rate of cycling is different across individuals. In our model,  $\omega_A$  is updated every 50 rounds, and it is sampled from an exponential distribution with mean  $\omega'_A$ , which remains constant for an agent.

In bipolar agents, the perceived reward is dependent on the agent's mood<sup>26,102</sup>, and hence cannot be modeled by a constant reward sensitivity. Instead, a time-varying reward sensitivity,  $\beta_A^t$  is used to model bipolarity.

$$\beta_A^t = \begin{cases} 1 & v_A^t = 0 \\ v_A^t + 1 & -1 \leq v_A^t < 0 \\ 1 + v_A^t(\beta_{\max} - 1) & 0 < v_A^t \leq 1 \end{cases} \quad (13)$$

The emotion induced by an interaction for bipolar agents is given by Eq. (10). The only difference is that, the reward sensitivity  $\beta_A^t$  is constant for manic and depressed agents, but changes with time for bipolar agents. It is clear from Eq. (13), that the agent overestimates reward when manic, and perceives diminished rewards when depressed.

**Social interactions and mood disorders.** Another symptom associated with manic and depressed phases is the relatively higher and lower drive of the individual. We already alluded to this concept in "Interaction model", stating agents can choose if they want to participate or not in any round of interaction. This is modeled as drive of an agent  $\zeta_A$ , with  $0 \leq \zeta_A \leq 1$ . In each round, the agent decides to participate with probability  $\zeta_A$ , given by:

$$\zeta_A = \begin{cases} 0.5 & \beta_A^0 = 1 \\ 0.5\beta_A^0 & 0 \leq \beta_A^0 < 1 \\ 0.5 + 0.5 \frac{\beta_A^0 - 1}{\beta_{\max} - 1} & 1 < \beta_A^0 \leq \beta_{\max} \end{cases} \quad (14)$$

The size of  $A$ 's neighborhood,  $\eta_A$  is also vastly different for manic and depressed agents with depressed agents being more lonely and having smaller personal network<sup>50,91</sup>. The social reach is a function of reward sensitivity, and is given by:

$$\eta_A = \begin{cases} 0.5|\mathcal{A}| & \beta_A^0 = 1 \\ 0.5\beta_A^0|\mathcal{A}| & 0 \leq \beta_A^0 < 1 \\ \left(0.5 + 0.5 \frac{\beta_A^0 - 1}{\beta_{\max} - 1}\right)|\mathcal{A}| & 1 < \beta_A^0 < \beta_{\max} \end{cases} \quad (15)$$

Both these parameters are constant for an agent, and Eqs. (14) and (15) are obtained by applying min–max rescaling to transform the ranges. For example, in the case of depressed agents,  $0 \leq \beta_A^0 \leq 1$  and  $0 \leq \zeta_A < 0.5$ . Hence, min–max rescaling is used to transform the value of  $\beta_A^0$  to an appropriate value of  $\zeta_A$ .

The mood is a complex phenomenon<sup>103</sup>. Also, there is no direct relationship between mood and behavior<sup>104</sup>. It is fuzzy, ambiguous, could be fleeting or lasting, and a person cannot pinpoint the "why" of their current mood<sup>105</sup>. There is also no direct causal relationship between mood and the drive to participate in an activity<sup>105</sup>. Therefore, there is no precise way to model drive and network size as state-like factors dependent on current mood.

## Agent types and design

An agent society comprising four different types of agents is considered. The agents are categorized into these types by behavior, decision making and internal attributes. The system is initialized with various specifications such as the number of agents, proportion of various types, and the number of rounds of interaction. The set of all agents is denoted by  $\mathcal{A}$  and the number of rounds is denoted by  $K$ . In the following subsections, we will define the attributes of an agent and the different types of agents.

**Agent attributes.** An agent  $A$  has several attributes that determine its nature and behavior. First,  $A$  has a unique identifier, denoted by  $\alpha_A$ . The total number of agents in the system is  $|\mathcal{A}|$ , therefore  $\alpha_A \in \{1, 2, \dots, |\mathcal{A}|\}$ . The performances of agents are compared based on cumulative payoffs at the end of the simulation, and this is captured by a non-negative real number. The mood of  $A$ ,  $v_A^t$  and the rate of change of mood  $\kappa$ , are as described in "Mood and memory". The reward sensitivity  $\beta_A^t$  is determined by its type, as discussed in "Mood disorders". The size of  $A$ 's neighborhood  $\eta_A$  is determined by the initial reward sensitivity  $\beta_A^0$  as given by Eq. (15). The drive of  $A$   $\zeta_A$ , a measure of the level of engagement, is also a function of  $\beta_A^0$  and is given by Eq. (14).  $A$  also maintains a set of its neighbors,  $N_A$ .

The memory of  $A$  has five attributes. The rate of forgetting  $\theta_A$  and the minimum threshold of trace strength  $\phi_{\min}$ , are as in "Forgetting". Another attribute is the level of mood dependence used to model mood-dependent retrieval in Eq. (9). It also needs to maintain a pair of gradients ( $k_N, k_P$ ) which are used as positive and negative gradients in the encoding function (Eq. 5). Lastly, the memory maintains a set of traces.

Each trace contains information pertaining to a single interaction, such as the identifier of the counterpart agent involved and the level of cooperation extended. Other information include the emotion  $\delta_A^t$  induced by the interaction (as given by Eqs. (3) and (10)). The trace is encoded and assigned an initial strength  $\phi_0$  (as given by Eq. (5)) and this is also captured. The strength of the trace,  $\phi_t$  decays with time (as given by Eq. (7)).

**Agent types.** As discussed in "Model of agent psychology and interactions", we consider four different types of agents: rational, manic, depressed, and bipolar. Table 1 summarizes the key differences among them.

Attribute	Rational	Manic	Depressed	Bipolar
$\beta_A^0$	1	[1,10]	[0,1]	[0,10]
$v_A^0$	0	[0,1]	[-1,0]	[-1,1]
$\kappa_A$	0	0.1	0.1	N/A
$\eta_A$	0.5  $\infty$	[0.5  $\infty$  ,   $\infty$  ]	[0, 0.5  $\infty$  ]	[0,   $\infty$  ]
$\zeta_A$	0.5	[0.5,1]	[0,0.5]	[0,1]

**Table 1.** Summary of agent types and attributes.

**Rational agents.** A rational agent is considered the baseline. It perceives rewards as they are, hence has a reward sensitivity of exactly 1, and its mood always stays neutral at 0. The size of its neighborhood is exactly half the total number of agents in the system (Eq. 15). Similarly, its drive is also 0.5, based on Eq. (14). This agent type is based upon the ideal of a perfect rational human.

**Manic agents.** A manic agent is the virtual equivalent of a human with chronic mania. For simulation, the value of  $\beta_{\max}$  is taken to be 10 and hence, the reward sensitivity lies between 1 and 10 (see "Mood disorders"). Mood  $v_A^t$  is initialized with a positive value based on  $\beta_A$  by Eq. (11), and the mood always remains positive. Manic agents exhibit greater drive than rational agents, and therefore  $\zeta_A > 0.5$  by Eq. (14). Manic agents' higher levels of social engagement are modeled with larger neighborhoods, from Eq. (15).

**Depressed agents.** A depressed agent is based on humans with Major Depressive Disorder. The reward sensitivity of these agents lies between 0 and 1 (see "Mood disorders"). Mood  $v_A^t$  is initialized with a negative value based on  $\beta_A$ , by Eq. (11), and the mood always remains negative. Depressed agents possess smaller neighborhoods as given by Eq. (15). Their levels of engagement are also lower, modeled with a lower drive than rational agents in Eq. (14).

**Bipolar agents.** A bipolar agent is subject to oscillating periods of mania and depression. As discussed in "Mood disorders", mood updates are modeled by a sine function. Bipolar agents have two additional attributes apart from the basic attributes defined in "Agent attributes":  $\omega_A$ , which controls the rate of cycling, and the offset  $\lambda_A$ .

The sizes of the neighborhood and the drive of bipolar agents are given by Eqs. (15) and (14), respectively. Since mood updates happen intrinsically due to internal rhythms according to Eq. (12), the rate of mood change  $\kappa_A$  is not pertinent for bipolar agents.

A summary of all four agent types and the permissible ranges for their attributes is given in Table 1.

## Results

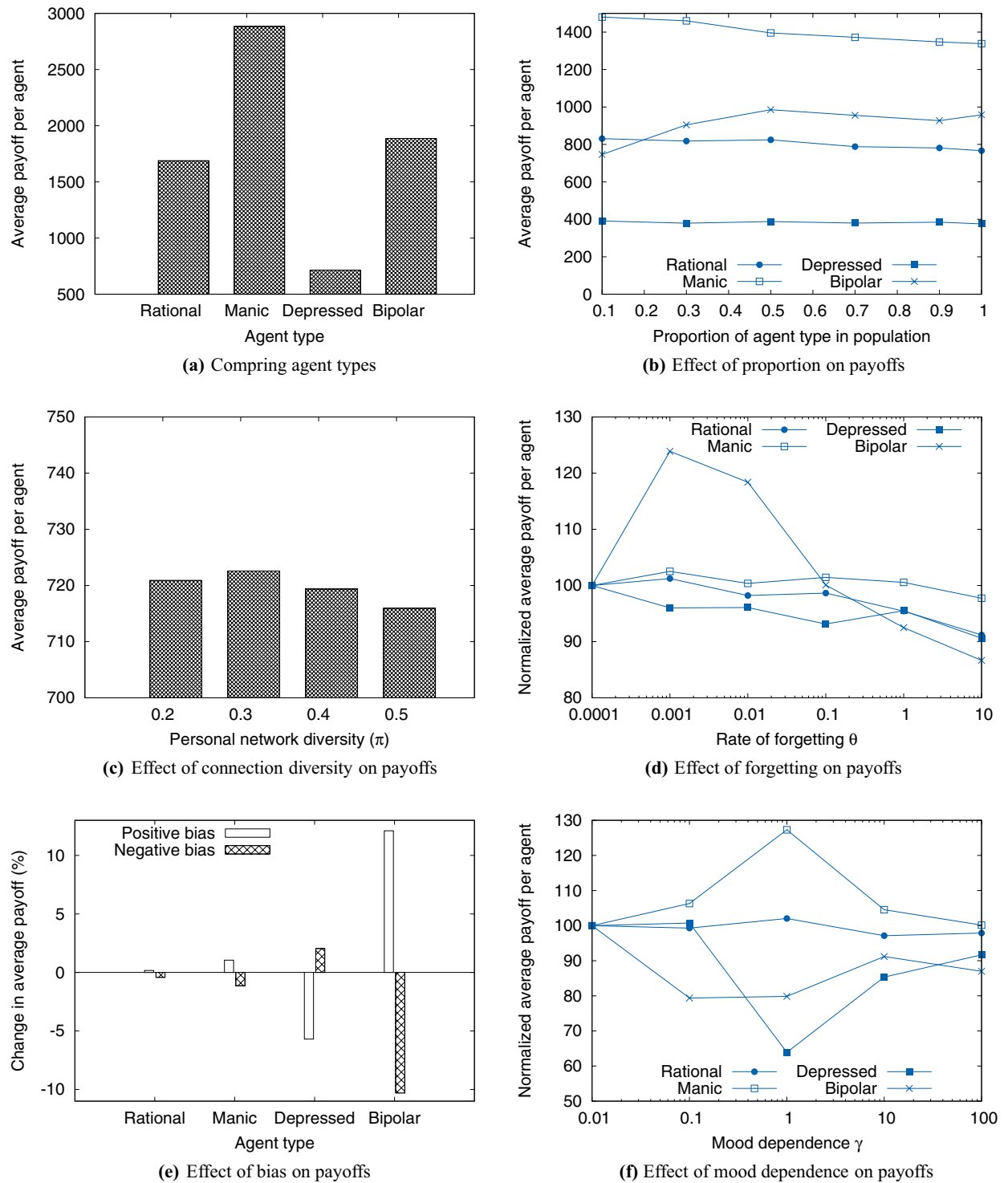
A system of 200 agents with representation from all four types is simulated for 2000 rounds of interaction. As discussed in "Agent types", the parameters of each agent are assigned in the range of permissible values. In each round, agents are paired according to the interaction model outlined in "Interaction model". The agents interact, update their payoffs, emotion, mood, and their availability for the next round. Different parameters of agents are varied in each experiment and their impacts on total payoff are presented in the subsequent subsections.

**Comparing payoffs of all agent types.** For this experiment, all four types of agents are equally represented (25%). We find that payoffs are highest for manic agents and least for depressed agents (Fig. 2a). Although bipolar agents show vastly different payoffs between themselves, on average the payoff of bipolar agents is greater than that of rational agents by about 10%. It is also seen that the average payoff of manic agents is about 70% higher than the average payoff of rational agents. When compared with rational agents, an average depressed agent's payoff is roughly 60% lower.

**Effect of varying the population composition.** In this experiment, for each type of agent, different levels of representation in the total population are considered, with the remaining types distributed equally. For example, if we consider 10% representation of rational agents in the system, then the remaining 90% is split between the other three types equally. In Fig. 2b, each curve depicts the effect on the average payoff of that type when the proportion of agents of that type is varied. No significant difference in the level of payoff is observed for rational and depressed agents when their proportion in the system is increased. However, increasing the proportion of manic agents resulted in a significant dip in their average payoff. For bipolar agents, a higher value of average payoff is observed with an increasing proportion of the corresponding type.

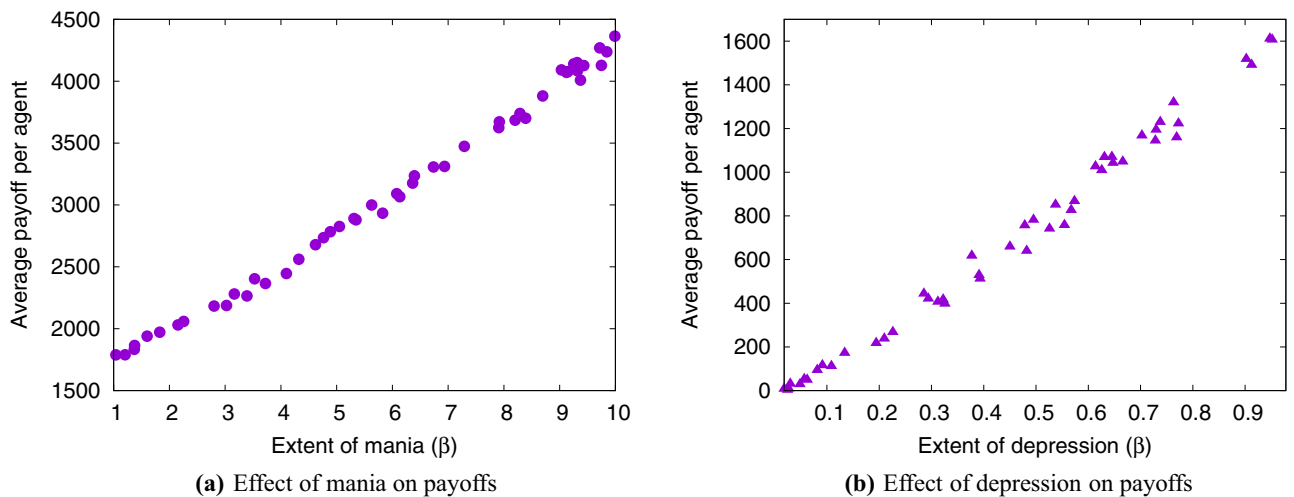
**Effect of diversity of connections in depressed agents.** In this experiment, we have an agent society with equal numbers of agent in all four types, and we vary the connection diversity of depressed agents from 0.2 to 0.5. A connection diversity of 0.25 implies there is an equal number of agents from all four types in the personal network of the agent. As seen in Table 1, the highest value of  $\eta$  for a depressed agent is 0.5| $\infty$ |. If  $\pi > 0.5$ , then the number of connections with depressed agents would exceed 0.25| $\infty$ |. This is not possible because the total number of agents of each type is 0.25| $\infty$ |. Through simulation, it is seen that a moderate value of  $\pi$  brings out the highest average payoff (Fig. 2c).





**Figure 2.** Results.

**Effect of forgetting.** In this experiment, the rate of forgetting is varied in powers of 10 starting from  $10^{-4}$  until 10. To avoid potential confounders, other parameters such as the positive and negative gradients, mood dependence, etc., are fixed. The value of gradients for the encoding function (Eq. 5),  $k_N$ ,  $k_P$  and the intercept  $c$  are chosen such that the initial strength of traces are less than 100. From Eq. (8), the minimum value of  $d$ , the duration of availability with  $\theta_A = 0.0001$  is much greater than 2000, which is the total number of interactions. Hence this corresponds to the case of a perfect memory, where an agent remembers every single interaction right from the beginning. On the other hand, when  $\theta_A = 10$ , the maximum duration of availability is less than 1, which implies the agent has no memory at all. Through simulation, it is observed that a stronger retention, i.e., a lower value of  $\theta_A$  offers better payoffs than weaker retention (Fig. 2d). However, the exact difference varies across types. Among manic agents, high values of  $\theta_A$  cause 4% lower payoffs than agents with perfect memory.



**Figure 3.** Effect of levels of mania and depression.

Among rational and depressed agents, those with no memory have an average payoff that is roughly 10% lower than agents with perfect memory. Among bipolar agents, agents with a near-perfect memory perform better than those with either a perfect memory or no memory.

**Effect of emotional bias.** Considering a system with equal proportions of all agent types, with other parameters such as  $\theta_A$  and  $\gamma_A$  fixed to avoid any potential confounders, the average payoff is considered the baseline (0), and the payoff associated with both biases are plotted as percentage differences from the baseline (Fig. 2e). In manic, and bipolar agents, positive bias is associated with higher payoffs, with positively biased agents performing about 12% better among bipolar agents, but only roughly 1% better among manic agents. There is not much change for rational agents. However, among depressed agents, negatively biased agents perform roughly 2% better than average.

**Effects of mood-based retrieval.** To analyze the effects mood dependence, we considered values from  $10^{-2}$  to 100 in powers of 10. For each type, the average payoff at the lowest level of mood dependence (0.01) is considered 100 and the payoffs at other levels are scaled appropriately. Fig. 2f shows that mood dependence does not have any significant effect on rational agents. However, for manic agents an intermediate value of mood dependence offers the highest payoff ( $\sim +30\%$ ) while in depressed agents, an intermediate value of mood dependence gives the lowest payoff ( $\sim -40\%$ ). In bipolar agents, low mood dependence maximizes payoff.

**Effect of reward sensitivity.** As discussed in "Mood disorders", the reward sensitivity  $\beta_A$  is a measure of the extent of mania or depression. Figure 3a depicts the almost-linear relationship between  $\beta_A$  and average payoff. It is observed that payoffs increase with increasing mania, which is fairly obvious because the higher the mania, higher is the drive in our model (Eq. 14). An almost-linear relationship between  $\beta_A$  and payoff is observed in depressed agents also (Fig. 3b). The observed relationship between payoffs and level of mania/depression is in line with prior clinical studies<sup>106,107</sup>, but there is no obvious analytical explanation in those. This model thus allows for insights into the observed outcomes. It is also clear that depressed agents at any level perform worse than manic agents.

## Conclusion

Mood disorders are unfortunately prevalent in society, hence models that capture them realistically are essential and relevant. We present a model that captures some important aspects of social interactions and success, given mood disorders and memory biases. A linear encoding function with different gradients for positive and negative emotions is used to model the relationship between emotion and memory. Mood-dependent memory retrieval is modeled using a triangular function centered at the agent's current mood. Social interactions are considered in mood disorders, as opposed to most prior work that considered agents in isolation.

Based on simulation of an agent society with different types of agents, we obtain results about relative payoffs and other aspects that are in agreement with, and extend, published studies. Our results concur with psychological studies that establish a relationship between severity of depression and lower performance<sup>106</sup>. Clinical studies also suggest diminished performance in depressed individuals and improved performance in cases of mania<sup>107–109</sup>, as also seen in our model.

Mood-dependent memory retrieval remains unexplored in prior works and clinical studies, and there seem to be no models that establish its impact on performance. We can show that an intermediate level of mood dependence brings the highest payoffs for manic agents, and lowest payoff for depressed agents. Our work can thus also be extended to effects of various traits or biases, some of which may not be easy to study empirically.

Received: 6 March 2020; Accepted: 21 October 2020

Published online: 01 December 2020

## References

- Schelling, T. C. *Micromotives and Macrobehavior* (Norton, Leicestershire, 1978).
- Epstein, J. M. & Axtell, R. *Growing artificial societies: Social science from the bottom up* (MIT Press, Cambridge, 1996).
- Harding, E. J., Paul, E. S. & Mendl, M. Cognitive bias and affective state. *Nature* **427**, 312. <https://doi.org/10.1038/427312a> (2004).
- Burman, O. H., Parker, R. M., Paul, E. S. & Mendl, M. T. Anxiety-induced cognitive bias in non-human animals. *Physiol. Behav.* **98**, 345–350 (2009).
- Sreenivas, N. K. & Rao, S. Egocentric bias and doubt in cognitive agents. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19*, 286–295 (International Foundation for Autonomous Agents and Multiagent Systems, 2019).
- Hegselmann, R. & Krause, U. Opinion dynamics and bounded confidence models, analysis and simulation. *J. Artif. Soc. Soc. Simul.* **5**, (2002).
- Sobkowicz, P. Opinion dynamics model based on cognitive biases of complex agents. *J. Artif. Soc. Soc. Simul.* <https://doi.org/10.18564/jasss.3867> (2018).
- Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. & Quattrocchi, W. Modeling confirmation bias and polarization. *Sci. Rep.* **7**, 40391. <https://doi.org/10.1038/srep40391> (2017).
- Lerner, J. S., Li, Y., Valdesolo, P. & Kassam, K. S. Emotion and decision making. *Annu. Rev. Psychol.* **66**, 799–823. <https://doi.org/10.1146/annurev-psych-010213-115043> (2015).
- Simon, H. A. *Reason in Human Affairs* (Stanford University Press, Stanford, 1983).
- Brown, R. & Kulik, J. Flashbulb memories. *Cognition* **5**, 73–99. [https://doi.org/10.1016/0010-0277\(77\)90018-X](https://doi.org/10.1016/0010-0277(77)90018-X) (1977).
- McCloskey, M., Wible, C. G. & Cohen, N. J. Is there a special flashbulb-memory mechanism?. *J. Exp. Psychol. Gen.* **117**, 171–181. <https://doi.org/10.1037/0096-3445.117.2.171> (1988).
- Walker, W. R., Vogl, R. J. & Thompson, C. P. Autobiographical memory: unpleasantness fades faster than pleasantness over time. *Appl. Cogn. Psychol.* **11**, 399–413. [https://doi.org/10.1002/\(SICI\)1099-0720\(199710\)11:5<399::AID-ACP462>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-0720(199710)11:5<399::AID-ACP462>3.0.CO;2-E) (1997).
- Rubin, D. C. & Berntsen, D. Life scripts help to maintain autobiographical memories of highly positive, but not highly negative, events. *Mem. Cogn.* **31**, 1–14. <https://doi.org/10.3758/BF03196077> (2003).
- Bluck, S. & Li, K. Z. Predicting memory completeness and accuracy: Emotion and exposure in repeated autobiographical recall. *Appl. Cogn. Psychol.* **15**, 145–158. [https://doi.org/10.1002/1099-0720\(200103/04\)15:2<145::AID-ACP693>3.0.CO;2-T](https://doi.org/10.1002/1099-0720(200103/04)15:2<145::AID-ACP693>3.0.CO;2-T) (2001).
- Arnold, M. B. *Emotion and personality* (Columbia University Press, New York, 1960).
- Lazarus, R. S. Progress on a cognitive-motivational-relational theory of emotion. *Am. Psychol.* **46**, 819–834. <https://doi.org/10.1037/0003-066X.46.8.819> (1991).
- Marinier, R. III. & Laird, J. Computational modeling of mood and feeling from emotion. *Proc. Annu. Meet. Cogn. Sci. Soc.* **29**, 461–466 (2007).
- Huys, Q. J., Pizzagalli, D. A., Bogdan, R. & Dayan, P. Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biol. Mood Anxiety Disord.* **3**, 12. <https://doi.org/10.1186/2045-5380-3-12> (2013).
- Lewis, P. A. & Critchley, H. D. Mood-dependent memory. *Trends Cogn. Sci.* **7**, 431–433. <https://doi.org/10.1016/j.tics.2003.08.005> (2003).
- Yates, J., Clare, L. & Woods, B. You've got a friend in me: Can Social Engagement mediate the relationship between mood and MCI?. *Innov. Aging* **1**, 1179. <https://doi.org/10.1093/geroni/igx004.4295> (2017).
- Brink, P. & Stones, M. Examination of the relationship among hearing impairment, linguistic communication, mood, and social engagement of residents in complex continuing-care facilities. *Gerontologist* **47**, 633–641. <https://doi.org/10.1038/427312a0> (2007).
- Shevchenko, Y. *The Influence of Mood on Decision-Making*. Ph.D. thesis, University of Mannheim (2018).
- Bagalman, E. & Cornell, A. S. Prevalence of Mental Illness in the United States: Data Sources and Estimates. *Congressional Research Service* **R43047**, (2018).
- Dayan, P. & Niv, Y. Reinforcement learning: The good, the bad and the ugly. *Curr. Opin. Neurobiol.* **18**, 185–196. <https://doi.org/10.1016/j.comb.2008.08.003> (2008) (**Cognitive neuroscience**).
- Eldar, E., Rutledge, R. B., Dolan, R. J. & Niv, Y. Mood as representation of momentum. *Trends Cogn. Sci.* **20**, 15–24. <https://doi.org/10.1016/j.tics.2015.07.010> (2016).
- Eldar, E. & Niv, Y. Interaction between emotional state and learning underlies mood instability. *Nat. Commun.* **6**, 6149. <https://doi.org/10.1038/ncomms7149> (2015).
- Clark, J. E., Watson, S. & Friston, K. J. What is mood? a computational perspective. *Psychol. Med.* **48**, 2277–2284. <https://doi.org/10.1017/S0033291718000430> (2018).
- Barger, S. D., Messerli-Bürge, N. & Barth, J. Social relationship correlates of major depressive disorder and depressive symptoms in Switzerland: nationally representative cross sectional study. *BMC Public Health* **14**, 273. <https://doi.org/10.1186/1471-2458-14-273> (2014).
- Teo, A. R., Choi, H. & Valenstein, M. Social relationships and depression: Ten-year follow-up from a nationally representative study. *PLoS One* **8**, e62396. <https://doi.org/10.1371/journal.pone.0062396> (2013).
- Kessler, R. *et al.* Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. *Am. J. Psychiatry* **163**, 1561–8. <https://doi.org/10.1176/appi.ajp.163.9.1561> (2006).
- Shippee, N. D. *et al.* Differences in demographic composition and in work, social, and functional limitations among the populations with unipolar depression and bipolar disorder: Results from a nationally representative sample. *Health Qual. Life Outcomes* **9**, 90. <https://doi.org/10.1186/1477-7525-9-90> (2011).
- Bonsall, M. B., Geddes, J. R., Goodwin, G. M. & Holmes, E. A. Bipolar disorder dynamics: Affective instabilities, relaxation oscillations and noise. *J. R. Soc. Interface* **12**, 20150670. <https://doi.org/10.1098/rsif.2015.0670> (2015).
- Goldbeter, A. Origin of cyclicity in bipolar disorders: A computational approach. *Pharmacopsychiatry* **46**, S44–S52. <https://doi.org/10.1055/s-0033-1341502> (2013).
- Lopez, A. *Markov Models for Longitudinal Course of Youth Bipolar Disorder*. Ph.D. thesis, University of Pittsburgh (2008).
- Urošević, S., Abramson, L. Y., Harmon-Jones, E. & Alloy, L. B. Dysregulation of the behavioral approach system (BAS) in bipolar spectrum disorders: Review of theory and evidence. *Clin. Psychol. Rev.* **28**, 1188–1205. <https://doi.org/10.1016/j.cpr.2008.04.004> (2008).
- Alloy, L. B. & Abramson, L. Y. The role of the behavioral approach system (BAS) in bipolar spectrum disorders. *Curr. Dir. Psychol. Sci.* **19**, 189–194 (2010).
- Rolls, E. T. *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach* (Oxford University Press, Oxford, 2007).
- Dudai, Y. How big is human memory, or on being just useful enough. *Learn. Mem.* **3**, 341–365. <https://doi.org/10.1101/lm.3.5.341> (1997).

40. Kandel, E. R., Dudai, Y. & Mayford, M. R. The molecular and systems biology of memory. *Cell* **157**, 163–186. <https://doi.org/10.1016/j.cell.2014.03.001> (2014).
41. Deese, J. & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. Exp. Psychol.* **54**, 180–187. <https://doi.org/10.1037/h0040536> (1957).
42. Murdock, B. B. Jr. The serial position effect of free recall. *J. Exp. Psychol.* **64**, 482–488. <https://doi.org/10.1037/h0045106> (1962).
43. Richards, B. A. & Frankland, P. W. The persistence and transience of memory. *Neuron* **94**, 1071–1084. <https://doi.org/10.1176/appi.ajp.163.9.15610> (2017).
44. Wozniak, P. A., Gorzelanczyk, E. J. & Murakowski, J. A. Two components of long term memory. *Acta Neurobiol. Exp.* **55**, 301–305 (1995).
45. Rubin, D. C., Hinton, S. & Wenzel, A. The precise time course of retention. *J. Exp. Psychol. Learn. Mem. Cogn.* **25**, 1161–1176. <https://doi.org/10.1176/appi.ajp.163.9.15611> (1999).
46. White, K. G. Forgetting functions. *Anim. Learn. Behav.* **29**, 193–207. <https://doi.org/10.1176/appi.ajp.163.9.15612> (2001).
47. Conte, R. & Paolucci, M. On agent-based modeling and computational social science. *Front. Psychol.* **5**, 668. <https://doi.org/10.1176/appi.ajp.163.9.15613> (2014).
48. Jackson, J. C., Rand, D., Lewis, K., Norton, M. I. & Gray, K. Agent-based modeling: A guide for social psychologists. *Soc. Psychol. Pers. Sci.* **8**, 387–395. <https://doi.org/10.1176/appi.ajp.163.9.15614> (2017).
49. Eberlen, J., Scholz, G. & Gagliolo, M. Simulate this! An introduction to agent-based models and their power to improve your research practice. *Int. Rev. Soc. Psychol.* **30**, 149–160. <https://doi.org/10.1176/appi.ajp.163.9.15615> (2017).
50. Fletcher, K., Parker, G., Paterson, A. & Synnott, H. High-risk behaviour in hypomanic states. *J. Affect. Disord.* **150**, 50–56. <https://doi.org/10.1176/appi.ajp.163.9.15616> (2013).
51. Rizvi, S. J., Pizzagalli, D. A., Sproule, B. A. & Kennedy, S. H. Assessing anhedonia in depression: Potentials and pitfalls. *Neurosci. Biobehav. Rev.* **65**, 21–35. <https://doi.org/10.1176/appi.ajp.163.9.15617> (2016).
52. Ecker, U. & Lewandowsky, S. Computational constraints in cognitive theories of forgetting. *Front. Psychol.* **3**, 400. <https://doi.org/10.3389/fpsyg.2012.00400> (2012).
53. Wixted, J. T. The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.* **55**, 235–269. <https://doi.org/10.1176/appi.ajp.163.9.15619> (2004).
54. Oberauer, K. & Lewandowsky, S. Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference?. *Psychol. Rev.* **115**, 544–76. <https://doi.org/10.1186/1477-7525-9-900> (2008).
55. Baddeley, A. Working memory. *Science* **255**, 556–559. <https://doi.org/10.1186/1477-7525-9-901> (1992).
56. Baddeley, A. D. & Hitch, G. Working memory. vol. 8 of *Psychology of Learning and Motivation*, 47 – 89, [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1) (Academic Press, Cambridge, 1974).
57. Page, M. P. A. & Norris, D. The primacy model: A new model of immediate serial recall. *Psychol. Rev.* **105**, 761–781. <https://doi.org/10.1186/1477-7525-9-902> (1998).
58. Burgess, N. & Hitch, G. Memory for serial order: A network model of the phonological loop and its timing. *Psychol. Rev.* **106**, 551–581. <https://doi.org/10.1186/1477-7525-9-903> (1999).
59. Lewandowsky, S. & Farrell, S. Short-term memory: New data and a model. In Ross, B. H. (ed.) *Advances in Research and Theory*, vol. 49 of *Psychology of Learning and Motivation*, 1 – 48, [https://doi.org/10.1016/S0079-7421\(08\)00001-7](https://doi.org/10.1016/S0079-7421(08)00001-7) (Academic Press, Cambridge, 2008).
60. Rubin, D. C. & Wenzel, A. E. One hundred years of forgetting: A quantitative description of retention. *Psychol. Rev.* **103**, 734–760. <https://doi.org/10.1186/1477-7525-9-904> (1996).
61. Wixted, J. T. & Ebbesen, E. B. Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Mem. Cogn.* **25**, 731–739. <https://doi.org/10.3758/BF03211316> (1997).
62. Wixted, J. T. & Ebbesen, E. B. On the form of forgetting. *Psychol. Sci.* **2**, 409–415 (1991).
63. Kahana, M. J. & Adler, M. Note on the power law of forgetting. *bioRxiv* <https://doi.org/10.1101/173765> (2002).
64. Levine, L. J. & Pizarro, D. A. Emotion and memory research: A grumpy overview. *Soc. Cogn.* **22**, 530–554. <https://doi.org/10.1521/soco.22.5.530.50767> (2004).
65. Burke, A., Heuer, F. & Reisberg, D. Remembering emotional events. *Mem. Cogn.* **20**, 277–290. <https://doi.org/10.3758/BF03199665> (1992).
66. Berntsen, D. Tunnel memories for autobiographical events: Central details are remembered more frequently from shocking than from happy experiences. *Mem. Cogn.* **30**, 1010–1020. <https://doi.org/10.3758/BF03194319> (2002).
67. Rozin, P. & Royzman, E. B. Negativity bias, negativity dominance, and contagion. *Pers. Soc. Psychol. Rev.* **5**, 296–320. [https://doi.org/10.1207/S15327957PSPR0504\\_2](https://doi.org/10.1207/S15327957PSPR0504_2) (2001).
68. Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. Bad is stronger than good. *Rev. Gen. Psychol.* **5**, 323–370. <https://doi.org/10.1037/1089-2680.5.4.323> (2001).
69. Klein, J. G. Negativity in impressions of presidential candidates revisited: The 1992 election. *Pers. Soc. Psychol. Bull.* **22**, 288–295. <https://doi.org/10.1098/rsif.2015.06701> (1996).
70. Ito, T. A., Larsen, J. T., Smith, N. K. & Cacioppo, J. T. Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *J. Pers. Soc. Psychol.* **75**(4), 887–900 (1998).
71. Skowronski, J. J. & Carlston, D. E. Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *J. Pers. Soc. Psychol.* **52**, 689–699. <https://doi.org/10.1098/rsif.2015.06702> (1987).
72. Dreben, E. K., Fiske, S. T. & Hastie, R. The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation. *J. Pers. Soc. Psychol.* **37**, 1758–1768. <https://doi.org/10.1098/rsif.2015.06703> (1979).
73. Bower, G. H. Mood and memory. *Am. Psychol.* **36**, 129–148. <https://doi.org/10.1098/rsif.2015.06704> (1981).
74. Snyder, M. & White, P. Moods and memories: Elation, depression, and the remembering of the events of one's life. *J. Pers.* **50**, 149–167. <https://doi.org/10.1111/j.1467-6494.1982.tb01020.x> (1982).
75. Erk, S. *et al.* Emotional context modulates subsequent memory effect. *NeuroImage* **18**, 439–447. <https://doi.org/10.1098/rsif.2015.06705> (2003).
76. Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. & Cahill, L. Event-related activation in the human amygdala associates with later memory for individual emotional experience. *J. Neurosci.* **20**, RC99. <https://doi.org/10.1523/JNEUROSCI.20-19-j0004.2000> (2000).
77. Drace, S. & Desrichard, O. Mood congruence effect in autobiographical recall: Is mood a mediator?. *Psihologija* **46**, 217–228 (2013).
78. Cochran, A. L., Schultz, A., McInnis, M. G. & Forger, D. B. *A Comparison of Mathematical Models of Mood in Bipolar Disorder*, 315–341 (Springer International Publishing, Cham, 2017).
79. Grant, B. *et al.* Prevalence, correlates, and comorbidity of bipolar i disorder and axis i and ii disorders: Results from the national epidemiologic survey on alcohol and related conditions. *J. Clin. Psychiatry* **66**, 1205–15. <https://doi.org/10.4088/JCP.v66n1001> (2005).
80. Cochran, A. L., Schultz, A., McInnis, M. G. & Forger, D. B. Testing frameworks for personalizing bipolar disorder. *Transl. Psychiatry* **8**, 36. <https://doi.org/10.1038/s41398-017-0084-4> (2018).
81. Daugherty, D. *et al.* Mathematical models of bipolar disorder. *Commun. Nonlinear Sci. Numer. Simul.* **14**, 2897–2908. <https://doi.org/10.1098/rsif.2015.06708> (2009).

82. Barrios, C., Goodnick, P. J. & Chaudhry, T. A. Rapid cycling bipolar disorder. *Expert Opin. Pharmacother.* **2**, 1963–1973. <https://doi.org/10.1098/rsif.2015.06709> (2001).
83. Coryell, W. *et al.* The long-term course of rapid-cycling bipolar disorder. *Arch. Gen. Psychiatry* **60**, 914–920. <https://doi.org/10.1001/archpsyc.60.9.914> (2003).
84. Steinacher, A. & Wright, K. A. Relating the bipolar spectrum to dysregulation of behavioural activation: A perspective from dynamical modelling. *PLoS One* **8**, 1–10. <https://doi.org/10.1371/journal.pone.0063345> (2013).
85. Chang, S.-S. & Chou, T. A dynamical bifurcation model of bipolar disorder based on learned expectation and asymmetry in mood sensitivity. *Comput. Psychiatry* **2**, 205–222. [https://doi.org/10.1162/cpsy\\_a\\_00021](https://doi.org/10.1162/cpsy_a_00021) (2018).
86. Murray, G. & Harvey, A. Circadian rhythms and sleep in bipolar disorder. *Bipolar Disord.* **12**, 459–472. <https://doi.org/10.1055/s-0033-13415023> (2010).
87. Kaladchibachi, S. A., Doble, B., Anthopoulos, N., Woodgett, J. R. & Manoukian, A. S. Glycogen synthase kinase 3, circadian rhythms, and bipolar disorder: A molecular link in the therapeutic action of lithium. *J. Circ. Rhythms* **5**, 3. <https://doi.org/10.1186/1740-3391-5-3> (2007).
88. Myin-Germeys, I. *et al.* Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr. Scand.* **107**, 124–131. <https://doi.org/10.1055/s-0033-13415025> (2003).
89. Pizzagalli, D. A., Jahn, A. L. & O’Shea, J. P. Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biol. Psychiatry* **57**, 319–327. <https://doi.org/10.1016/j.biopsych.2004.11.026> (2005).
90. Berrios, G. E. Of mania: Introduction. *Hist. Psychiatry* **15**, 105–124. <https://doi.org/10.1177/0957154X04041829> (2004).
91. Romans, S. E. & McPherson, H. M. The social networks of bipolar affective disorder patients. *J. Affect. Disord.* **25**, 221–228. <https://doi.org/10.1055/s-0033-13415027> (1992).
92. McKenzie, L. E. *et al.* Social contacts and depression in middle and advanced adulthood: Findings from a us national survey, 2005–2008. *Int. J. Soc. Psychiatry* **59**, 627–635. <https://doi.org/10.1055/s-0033-13415028> (2013).
93. Acevedo, M. & Krueger, J. I. Evidential reasoning in the prisoner’s dilemma. *Am. J. Psychol.* **118**, 431–457 (2005).
94. Ng, G. T. T. & Au, W. T. Expectation and cooperation in prisoner’s dilemmas: The moderating role of game riskiness. *Psychon. Bull. Rev.* **23**, 353–360. <https://doi.org/10.3758/s13423-015-0911-7> (2016).
95. Verhoeff, T. *A continuous version of the prisoner’s dilemma*. Computing science notes (Technische Universiteit Eindhoven, 1993).
96. Killingback, T. & Doebeli, M. The continuous prisoner’s dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *Am. Nat.* **160**, 421–438 (2002).
97. Jaggi, N. & Vij, D. Fourier transform infrared spectroscopy. In *Handbook of Applied Solid State Spectroscopy* (ed. Vij, D. R.), 411–450. [https://doi.org/10.1007/0-387-37590-2\\_9](https://doi.org/10.1007/0-387-37590-2_9) (Springer, Boston, 2006).
98. Alloy, L. B., Olino, T., Freed, R. D. & Nusslock, R. Role of reward sensitivity and processing in major depressive and bipolar spectrum disorders. *Behav. Ther.* **47**, 600–621. <https://doi.org/10.1016/j.beth.2016.02.014> (2016).
99. Nusslock, R., Abramson, L. Y., Harmon-Jones, E., Alloy, L. B. & Hogan, M. E. A goal-striving life event and the onset of hypomanic and depressive episodes and symptoms: Perspective from the behavioral approach system (BAS) dysregulation theory. *J. Abnorm. Psychol.* **116**, 105–115. <https://doi.org/10.1037/0021-843X.116.1.105> (2007).
100. Rosenquist, J., Fowler, J. & Christakis, N. Social network determinants of depression. *Mol. Psychiatry* **16**, 273–81. <https://doi.org/10.1038/mp.2010.13> (2010).
101. Goldbeter, A. A model for the dynamics of bipolar disorders. *Prog. Biophys. Mol. Biol.* **105**, 119–127. <https://doi.org/10.1016/j.pbiomolbio.2010.11.007> (2011) (**BrainModes: The role of neuronal oscillations in health and disease**).
102. Bennett, D. & Niv, Y. Opening burton’s clock: Psychiatric insights from computational cognitive models. <https://doi.org/10.31234/osf.io/y2vzu> (2018).
103. Gilboa, E. & Revelle, W. Personality and the Structure of Affective Responses. In *Emotions: Essays on Emotion Theory* (eds. van Gozen, S. H. M., Van de Poll, N. E. & Sergeant, J. A.), chap. 5, 135–159 (Lawrence Erlbaum Associates, Inc., 1994).
104. Revelle, W. Personality and motivation: Sources of inefficiency in cognitive performance. *J. Res. Pers.* **21**, 436–452. <https://doi.org/10.1016/j.cpr.2008.04.0042> (1987).
105. Gendolla, G. H. E., Brinkmann, K. & Richter, M. Mood, motivation, and performance: An integrative theory, research, and applications. In *Mood and Human Performance: Conceptual, Measurement, and Applied Issues* (eds. Lane, A. M.), chap. 2, 35–61 (Nova Science Publisher, Inc., New York, 2007).
106. Kizilbash, A. H., Vanderploeg, R. D. & Curtiss, G. The effects of depression and anxiety on memory performance. *Arch. Clin. Neuropsychol.* **17**, 57–67. [https://doi.org/10.1016/S0887-6177\(00\)00101-3](https://doi.org/10.1016/S0887-6177(00)00101-3) (2002).
107. Koenen, K. C. *et al.* Childhood iq and adult mental disorders: A test of the cognitive reserve hypothesis. *Am. J. Psychiatry* **166**, 50–57. <https://doi.org/10.1016/j.cpr.2008.04.0044> (2009).
108. Cannon, M. *et al.* Evidence for early-childhood, pan-developmental impairment specific to schizophreniform disorder: Results from a longitudinal birth cohort. *Arch. Gen. Psychiatry* **59**, 449–456. <https://doi.org/10.1001/archpsyc.59.5.449> (2002).
109. Lerner, D. *et al.* Work performance of employees with depression: The impact of work stressors. *Am. J. Health Promot.* **24**, 205–213. <https://doi.org/10.4278/ajhp.090313-QUAN-103> (2010).

### Author contributions

N.S. and S.R. conceived the model and experiments, N.S. implemented the code and analysed the results. All authors reviewed the manuscript.

### Competing Interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020