



Brain-Inspired Affective Empathy Computational Model and Its Application on Altruistic Rescue Task

Hui Feng^{1,2}, Yi Zeng^{1,2,3,4*} and Enmeng Lu¹

¹ Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China,

² School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, ³ Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China, ⁴ National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:

Giovanni Mirabella,
University of Brescia, Italy

Reviewed by:

Andres Pinilla,
Technical University of Berlin,
Germany
Radwa Khalil,
Jacobs University Bremen, Germany
Ye Yuan,
Shanghai Jiao Tong University, China

*Correspondence:

Yi Zeng
yi.zeng@ia.ac.cn

Received: 11 February 2022

Accepted: 22 June 2022

Published: 18 July 2022

Citation:

Feng H, Zeng Y and Lu E (2022)
Brain-Inspired Affective Empathy
Computational Model and Its
Application on Altruistic Rescue Task.
Front. Comput. Neurosci. 16:784967.
doi: 10.3389/fncom.2022.784967

Affective empathy is an indispensable ability for humans and other species' harmonious social lives, motivating altruistic behavior, such as consolation and aid-giving. How to build an affective empathy computational model has attracted extensive attention in recent years. Most affective empathy models focus on the recognition and simulation of facial expressions or emotional speech of humans, namely Affective Computing. However, these studies lack the guidance of neural mechanisms of affective empathy. From a neuroscience perspective, affective empathy is formed gradually during the individual development process: experiencing own emotion—forming the corresponding Mirror Neuron System (MNS)—understanding the emotions of others through the mirror mechanism. Inspired by this neural mechanism, we constructed a brain-inspired affective empathy computational model, this model contains two submodels: (1) We designed an Artificial Pain Model inspired by the Free Energy Principle (FEP) to simulate the pain generation process in living organisms. (2) We build an affective empathy spiking neural network (AE-SNN) that simulates the mirror mechanism of MNS and has self-other differentiation ability. We apply the brain-inspired affective empathy computational model to the pain empathy and altruistic rescue task to achieve the rescue of companions by intelligent agents. To the best of our knowledge, our study is the first one to reproduce the emergence process of mirror neurons and anti-mirror neurons in the SNN field. Compared with traditional affective empathy computational models, our model is more biologically plausible, and it provides a new perspective for achieving artificial affective empathy, which has special potential for the social robots field in the future.

Keywords: affective empathy, mirror neuron system, spiking neural network, Artificial Pain, altruistic behavior, self-awareness

1. INTRODUCTION

Empathy is the ability to understand the state of others through observation, imagination, and inference, which is the motivation for altruistic behaviors such as consolation and aid-giving (Waal and Preston, 2017). Empathy plays a vital role in human society as it is the biological basis of social morality (Lou, 2011). In addition, empathy is cross-species (Malin et al., 2011), which maintains communication not only within the same species but also between different species. In general, empathy is fundamental to the harmonious coexistence of social groups.

Empathy is divided into cognitive empathy and affective empathy (Asada, 2015). Cognitive empathy refers to the ability that observers to imagine and infer the target's feeling or state. Affective empathy refers to the ability that observers are directly affected by the emotional state of another by matching it (Waal and Preston, 2017). Affective empathy is more fundamental and more common, and it appears in the early stage of life development and also exists in most non-human species such as rodents and birds (Shamay-Tsoory et al., 2009). Affective empathy also has essential research value in the field of social robotics as it is conducive to the communication and cooperation between robots and between humans and robots. Our work mainly focuses on the affective empathy computational model and its applications.

Emotion is an embodied internal state of the organism's brain that is formed through interaction with the environment (Lerner et al., 2015). Emotion not only represents the organism's state but also generates emotional overt actions such as facial expressions to communicate with peers and achieve affective empathy. From the psychology perspective, the observer can use the Perception-Action Mechanism (PAM) to empathize with others (Preston and de Waal, 2017). Perceiving other people's emotions will activate the same emotion representation in the observer's brain, which is equivalent to the observer experiencing that emotion. Recently, much neuroscience literature has shown that the Mirror Neuron System (MNS) supports the PAM (Rizzolatti and Sinigaglia, 2016). The MNS consists of a group of neurons with sensory-motor properties, and they are activated both during action execution and action observation (Erhan et al., 2006; Khalil et al., 2018b). The MNS a collection of Motor brain regions in the parietal and frontal lobes. In the process of affective empathy, when we perceive others' emotions, such as seeing their facial expressions or hearing them cry, we first achieve motor-level understanding through the MNS and then further achieve emotion-level understanding (Carr et al., 2003). It is worth noting that the activation patterns of the brain during experiencing own emotion and empathizing with others are different because of the presence of anti-mirror neurons (Christian and Valeria, 2010). Thus, the brain can distinguish who is the producer of emotion, which is called primary self-awareness (Lamm et al., 2011).

We argue that computational modeling of affective empathy should follow its neural mechanisms: First, the model should have an internal state similar to human emotion. Second, the model should develop its own mirror mechanism similar to the human MNS function, so as to understand others' emotions with its own experience. Third, the model should have the ability to distinguish self from others and adopt adaptive responding behaviors.

Most of the existing affective empathy computational models mainly belong to the Affective Computing field, which uses machine learning algorithms to recognize and simulate human emotion overt signals, such as facial expressions, speeches, and body movements (Claret et al., 2017; Soujanya et al., 2017; Huang et al., 2020; Lee and Kang, 2020). (Leite et al., 2014) designed an iCat robot to play chess with children. It can analyze the state of the chess game and the child's facial expressions to make a corresponding response such as encouragement.

Zheng et al. (2019) proposed a children's companion robot called BabeBay, which has affective computing ability for multi-modal information such as expression, body gesture, and text, to maintain companionship with different children. These traditional approaches are only computational processing of the emotion overt signals, but the emotion overt signals are not equal to the emotion. Emotion is a complex embodied internal state. It is obvious that these approaches do not define this internal state; they treat emotion as a mere label. Therefore, we believe that these approaches are far from enough to achieve real artificial affective empathy. In addition to that, these approaches also do not follow the guidance of any neural mechanism of affective empathy. Woo et al. (2017) implemented the affective empathy process using the spiking neural network. However, this study only used brain-like modeling tools and did not involve a mirror mechanism. Watanabe et al. (2007) modeled a communication model for the virtual robot similar to the 'intuitive parenting' process between babies and their caregivers. The robot can establish the relationship between the emotions and the caregiver's facial expressions by mirror mechanism, then they can respond to the other's emotion by observing human facial expressions. But again, this study did not define emotion as an internal state of the robot, and this model cannot distinguish the producer of emotion between self and others.

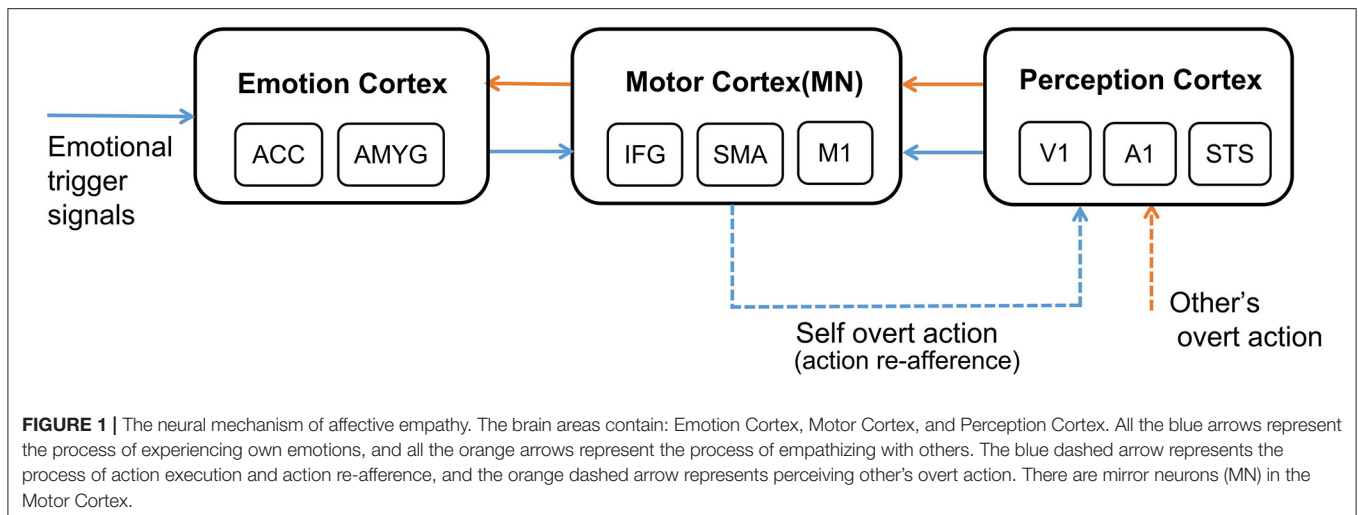
In this article, we construct a more sophisticated brain-inspired affective empathy computational model. First, we constructed a computational model of emotion as an internal state: the neural mechanisms by which emotions emerge have not been well studied. The Free Energy Principle (FEP) has been proposed as a unified Bayesian interpretation of perception, learning, and action, describing the relationship between the internal model of the brain and the relevant sensations from the environment (Joffily and Coricelli, 2013). Coincidentally, emotion is a neural activity formed through the interaction of brain expectations with the real situation of the environment (Brown and Brune, 2012). Therefore, in this article, we use FEP as a theoretical basis to model a specific emotion-pain. Second, we modeled the mirror mechanism of empathy: we used a spiking neural network to reproduce the emergence process of mirror neurons and anti-mirror neurons, the process of empathizing with others through mirror mechanism, and the ability of self-other differentiation.

The contributions of our study can be summarized as follows: (1) Inspired by FEP, we modeled the Artificial Pain as an internal state. (2) We constructed an affective empathy spiking neural network (AE-SNN), which can simulate the mirror mechanism of MNS to empathize with others and have the ability of self-others differentiation. (3) We also explored the intrinsic motivations of altruistic behavior, together with the above parts, to complete the pain empathy and altruistic rescue task in the grid world.

2. MATERIALS AND METHODS

2.1. The Neural Mechanism

The whole process of affective empathy is shown in **Figure 1**. It involves the following brain areas: On the far left is the Emotion Cortex. Different emotional trigger signals can result



in different firing patterns in this cortex, producing different emotional states. The Emotion Cortex contains many sub-regions, such as the Anterior Cingulate Cortex (ACC) and Amygdala (AMYG), which are essential components of the Limbic System (Reep et al., 2007). The ACC is often thought to be associated with pain (Corradi-Dell'Acqua et al., 2016). The AMYG is often thought to be associated with fear (Davis, 1992). In the middle is the Motor Cortex. In the affective empathy process, the Motor Cortex is mainly responsible for producing emotional overt actions (Mukamel et al., 2010), such as painful facial expressions, shouting, and crying. The Motor Cortex contains many sub-regions. Our study considers only three representative sub-regions: Inferior Frontal Gyrus (IFG), Supplementary Motor Cortex (SMA), and Primary Motor Cortex (M1). IFG is responsible for encoding the intention of actions (Jabbi, 2008). SMA is responsible for the initialization of the action sequence (Rizzolatti and Luppino, 2001). M1 guides the muscles to perform specific actions (Gazzola and Keysers, 2008). There are mirror neurons (MN) in the Motor Cortex (Erhan et al., 2013), as shown in **Figure 1**. On the far right is the Perception Cortex, which is used to perceive the emotional overt actions of oneself or others, such as seeing the painful facial expressions and hearing the cry. We also listed three of the most representative sub-regions in the Perception Cortex: Primary Auditory Cortex (A1) and Primary Visual Cortex (V1) perform primary processing of visual and auditory information (Zipser et al., 1996; Morosan et al., 2001). The Superior Temporal Sulcus (STS) is a high-order perception area that integrates visual and auditory information about the body and facial actions (Keysers and Gazzola, 2014). The connection between the Motor Cortex and the Emotion Cortex is bidirectional (Jabbi, 2008). The connection between the Perception Cortex and the Motor Cortex is also bidirectional (Kilner and Frith, 2007), but for the affective empathy process in our study, we only considered the connection from the Perception Cortex to the Motor Cortex.

Different emotional trigger signals will lead to the production of corresponding emotions. In this article, we explore a particular emotion–pain. External noxious stimuli specifically

activate the Nociceptors in the skin of organisms, then the noxious information will be transmitted to the cerebral cortex through the Anterolateral System (Asada, 2019). In the cerebral cortex, the pain will be evoked from two aspects: the sensory discrimination component will be formed in the Somatosensory Cortex (S1, S2), and the affective and motivational component will be formed in the Anterior Cingulate Cortex (ACC) (Donald, 2000). However, this pathway is specific and formed in thousands of years of evolution (Walters and Williams, 2019). We argue that the Artificial Pain Model should not follow the physiological mechanism completely, because many intelligent agents, such as robots, do not have a physiological structure similar to organisms. We should explore the essence and significance of pain for the survival of organisms.

Broom et al. proposed that the emergence of pain was first related to physical injury in the evolutionary process (Broom, 2001). Pain is a kind of neural activity that occurs with physical injury and is preserved and internalized in the brain because of its survival benefits. In further learning, the experience of pain can also be associated with injury-related cues, such as scenes or voices of injury. When a similar cue occurs again, the brain will produce the same pain experience and avoid potential injury (Wiech and Tracey, 2013).

Certain emotions will trigger the Motor Cortex to perform corresponding emotional overt action, as shown by the blue dashed arrow in **Figure 1**. After experiencing own emotions, the observer can use the mirror mechanism of the Mirror Neuron System (MNS) to empathize with others. Mirror neurons are widely present in the Motor Cortex, and both the previously mentioned IFG and SMA sub-regions contain mirror neurons (Rizzolatti and Sinigaglia, 2016). MNS is formed gradually during individual development. Keysers et al. proposed that the emergence of mirror neurons is due to synchronous action execution and action re-afference. You can see or hear your emotional overt action when you execute it. This kind of perceptual input resulting from your action is called action re-afference (Keysers and Gazzola, 2014), as shown by the blue

dashed arrow in **Figure 1**. The existence of the time overlap between action execution and action re-ference strengthens the synapse weights between the neurons in the Motor Cortex and the Perception Cortex representing the same action, and weakens the synapse weights representing different actions, resulting in some neurons of the Motor Cortex having mirror properties, namely mirror neurons. In the process of affective empathy, perceiving the emotional overt action of others will first activate the corresponding neurons in the Perception Cortex, as shown by the orange dashed arrow in **Figure 1**. Subsequently, the mirror neurons of the Motor Cortex that perform the same emotional overt action will fire, forming the internal motor level representation of other's emotion. The Motor Cortex is connected to the Emotion Cortex, which further activates the corresponding emotional neurons, forming an internal emotion-level representation of others' emotions, as shown by the orange solid arrow in **Figure 1**. For example, when a baby is experiencing emotions, it will instinctively activate the Motor Cortex to produce emotional overt actions (Yamada, 1993), such as crying or facial expressions. Babies can also perceive these emotional overt actions by their Auditory Cortex or Visual Cortex. When a baby performs a happy facial expression, their caregiver will imitate their expression so that they will also see the caregiver's expression, called Intuitive Parenting (Watanabe et al., 2007). Due to the similar activation time, the synaptic weights from the neurons in the Perception Cortex to the neurons in the Motor Cortex that represent the same emotional overt action will be strengthened, forming mirror neurons. When seeing others' facial expressions or hearing others' cries, they can understand others' emotions through the mirror mechanism.

In addition, the affective empathy network of the brain can distinguish the self from others. The reason is that the anti-mirror neurons in Motor Cortex are activated during action execution and deactivated during action observation (Christian and Valeria, 2010), leading to different firing patterns when executing own actions or only observing others' actions. The existence of anti-mirror neurons can distinguish who is the producer of the emotional overt action, which is also the embodiment of preliminary self-awareness. Anti-mirror neurons were found in the M1 region in Motor Cortex, the emergence of which is due to the gating mechanism of SMA neurons (Gazzola and Keysers, 2008). SMA neurons directly project to M1 neurons; the firing state of M1 neurons is totally decided by SMA neurons. SMA neurons will produce different firing patterns during action observation and action execution, which can control M1 neurons to activate during action execution and deactivate during action observation. This process is considered to be the basis for self-other differentiation (Mukamel et al., 2010).

2.2. The Architecture of the Computational Model

This subsection describes the overall architecture of the brain-inspired affective empathy computational model. As shown in **Figure 2**, the model is divided into two submodels: the Artificial Pain Model and the affective empathy spiking neural network (AE-SNN).

The Artificial Pain Model defines Artificial Pain from the Free Energy perspective, and it can receive different external stimuli and produce the corresponding pain state, the output of this model will be used as the input of the AE-SNN.

The AE-SNN simulates the function and connection of the brain regions mentioned in Section 2.1, which contains the Emotion Module, Motor Module, and Perception Module. Each module is a neuron population. The different emotional states are represented by population coding in the Emotion Module (Fang et al., 2021). The Motor Module represents the different emotional overt actions. The mirror neurons and anti-mirror neurons will emerge in Motor Module. Because anti-mirror neurons are formed due to the SMA neurons and M1 neurons, we set a certain number of neurons in this population to simulate the properties of M1 and SMA neurons. M1 neurons are only activated by SMA neurons, as shown by the orange curved arrow in **Figure 2**. The Perception Module represents the perception information of the different emotional overt actions, which are also encoded by population coding. The Motor Module and the Emotion Module are connected in a bidirectional way, all of which are excitatory connections. The connection from the Perception Module to the Motor Module contains inhibitory connections to SMA neurons and excitatory connections to other neurons. In **Figure 2**, the orange arrows indicate excitatory connections, and the blue arrow indicates inhibitory connections.

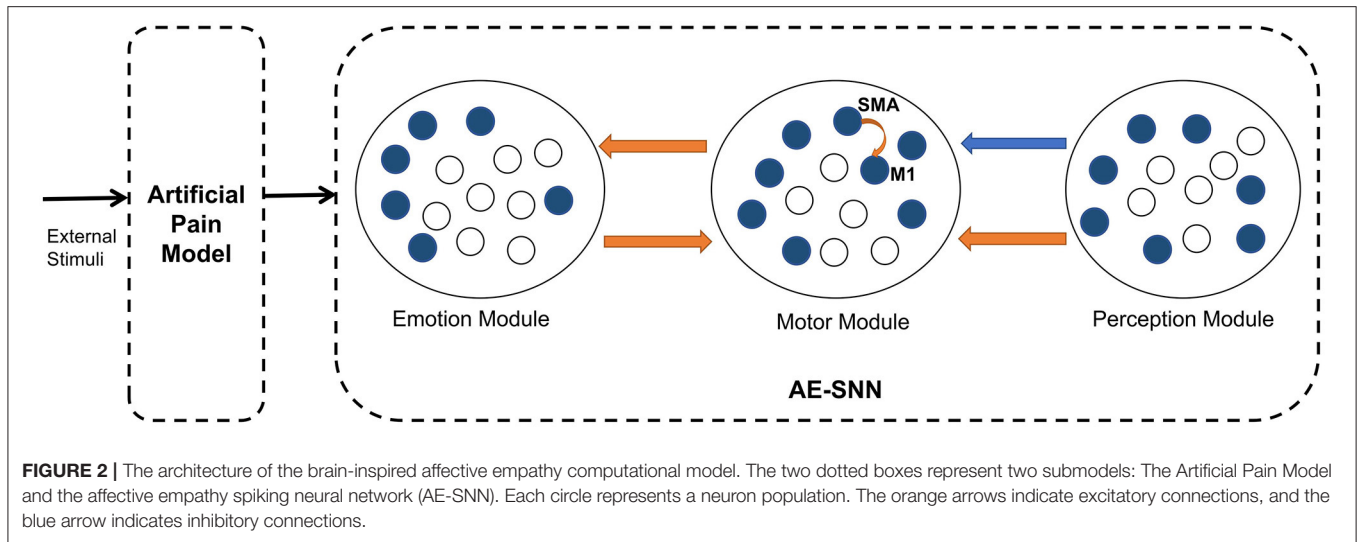
2.3. Model Implementation

2.3.1. Artificial Pain as an Internal State

We propose that the traditional approaches followed by previous studies in the Affective Computing field are far from enough to achieve real artificial affective empathy. Emotion should be modeled as the bottom-up internal state of the agent so that it can process environmental information more adaptively and carry out affective empathy more reasonably. This subsection discusses the modeling of a particular emotion-pain.

As mentioned in Subsection 2.1, the first step of the Artificial Pain Model is to quantify the actual damage of intelligent agents. From the perspective of Free Energy, body damage is an unexpected state, called the entropy increasing state (Friston, 2010). However, the entropy of organisms cannot be directly quantified; the Free Energy is the upper bound of entropy, and minimizing Free Energy approximates minimizing the entropy (Ramstead et al., 2018). Thus, FEP is a way to quantify the damage.

Bogacz demonstrated the relationship between Free Energy and entropy (Bogacz, 2017), providing the definition and expression of Free Energy in detail: the brain could not directly know the actual state of the external world ϕ ; it can only constantly estimate the state of the external world $\hat{\phi}$; the sensory information s received by the sense organs is used to verify the estimation. The Free Energy is finally simplified and expressed as the negative logarithm of the joint probability distribution of the state estimation $\hat{\phi}$ and the sensory information s . According to the definition, the larger the joint probability distribution, the smaller the Free Energy, and the more accurate the estimation.



Expanded by the Bayesian formula $p(s, \hat{\phi}) = p(s|\hat{\phi})p(\hat{\phi})$:

$$\begin{aligned}
 FE &= -\ln p(s, \hat{\phi}) \\
 &= -\ln p(s|\hat{\phi}) - \ln p(\hat{\phi})
 \end{aligned} \tag{1}$$

We replace the probability distribution with a Gaussian distribution $f(s; g(\cdot), \sigma)$ with mean $g(\cdot)$ and variance σ , supposing $\sigma = 1$:

$$\begin{aligned}
 FE &= -\ln f_1(s; g_s(\hat{\phi}), \sigma_s) - \ln f_2(\hat{\phi}; \mu_\phi, \sigma_\phi) \\
 &= \frac{[s - g_s(\hat{\phi})]^2}{2\sigma_s} + \frac{[\hat{\phi} - \mu_\phi]^2}{2\sigma_\phi} - \frac{1}{2} \ln \sigma_s - \frac{1}{2} \ln \sigma_\phi \tag{2} \\
 &= \frac{1}{2}[s - g_s(\hat{\phi})]^2 + \frac{1}{2}[\hat{\phi} - \mu_\phi]^2
 \end{aligned}$$

The first item in Equation 2 is the sensory dynamic (McGregor et al., 2015). $g_s(\cdot)$ is the sensory generating function, mapping the relationship between the world state and the sensory information learned by experience in advance. The brain actively estimates the world state $\hat{\phi}$, then predicts sensory information through the generating function $g_s(\hat{\phi})$. The actual sensory information s received through the sense organs is used to calculate prediction errors. The second term in Equation 2 is the environmental dynamic (McGregor et al., 2015). The estimation of the brain depends on prior knowledge μ_ϕ of the environment and this item is similar to the prior error of Bayes theorem. Environmental dynamic and sensory dynamic jointly determine the value of Free Energy.

For robotic systems, the world state mentioned above is the body state of the robot and the sensory information for robots are vision s_v and proprioception s_p (Lanillos and Cheng, 2018), as shown in Equation 3. The robot continuously estimates its body state $\hat{\phi}$, and predicts the vision $g_{sv}(\hat{\phi})$ and proprioception $g_{sp}(\hat{\phi})$, respectively. The generating functions $g_{sv}(\cdot)$ and $g_{sp}(\cdot)$ can be learned by motor babbling in advance (Lanillos and Cheng,

2018), which is the mapping relationship between the body state and the vision information or proprioception information. Therefore, the sensory dynamic of the robotic system includes visual prediction error and proprioception prediction error. For the environmental dynamic, the body state ϕ of the robot at the current moment is determined by the state ϕ and the action a' at the last moment. $g_\phi(\cdot)$ represents the mapping relationship. When the robot's body structure is deformed, or the sensor is damaged, the sensory dynamic will be affected. When the robot motor structure is damaged, the environmental dynamic will be affected. In these cases, the value of Free Energy will change. So the value of Free Energy is taken to evaluate the robot's body damage.

$$FE = \frac{1}{2}[s_v - g_{sv}(\hat{\phi})]^2 + \frac{1}{2}[s_p - g_{sp}(\hat{\phi})]^2 + \frac{1}{2}[\hat{\phi} - g_\phi(\phi', a')]^2 \tag{3}$$

In living organisms, physical damage can result in a pain state, and physical damage can be quantified by the value of Free Energy. Thus, in a robotic or virtual agent system, Artificial Pain can be considered relative to the value of Free Energy that represents their body damage. When the Free Energy is greater than 0, the body is in an injured state and then causes a pain state; When the Free Energy equals 0, the body is in a normal state.

In some potential dangerous scenarios, substantial injury has not occurred, but the organisms can also feel pain. This is because the brain can associate the experience of pain with injury-related cues so that it can quickly avoid similar cues. In the actual application of the robot, after the robot suffered substantial body damage, the robot's vision system should capture the corresponding dangerous scenarios. The robot will then associate it with the internal pain state and avoid potential damage. The significance of pain is that it is a warning signal for the actual and potential body damage of the living organisms to protect themselves (Melzack, 2001). The Artificial Pain created with the proposed model can not only generate a pain warning signal when the body damage actually occurs, but also avoid potential

damage in the future, achieving the same significance as the living organisms' pain.

2.3.2. The Affective Empathy Spiking Neural Network

When different internal emotional states are generated through the Artificial Pain Model, the AE-SNN simulating mirror mechanism will be trained. This subsection describes the concrete design and implementation of the AE-SNN.

1. LIF neuron model and STDP learning rule. Compared to conventional artificial neural networks, spiking neural network (SNN) captures more essential characteristics of the biological brain (Ghosh-Dastidar and Adeli, 2009; Khalil et al., 2017, 2018a; Zhao et al., 2018). In recent years, SNN has been successfully applied in many aspects of cognitive function modeling, such as decision making and creative processes (Héricé et al., 2016; Khalil and Moustafa, 2022). The Leaky Integrate-and-fire (LIF) neuron is a well accepted computational model for SNN neurons (Tal and Schwartz, 1997), and we use it as the basic unit of our model. The LIF neuron dynamics are described by the following Equation 4 and Equation 5:

$$\tau_m \frac{du}{dt} = -[u_t - u_{rest}] + RI(t) \quad (4)$$

$$\lim_{\delta \rightarrow 0} u(t^f + \delta) = u_{reset} \quad (5)$$

u_t is the membrane potential of the neuron at time t , u_{rest} is the membrane potential at steady-state, R is the resistance, $I(t)$ is the input current, and τ_m is the time constant. When the membrane potential u_t exceeds a certain threshold u_{th} , the neuron fires and t_f is the firing time. Once the neuron has fired, the membrane potential returns to its reset state u_{reset} (Fang et al., 2021). In this article, the parameters of the LIF neuron are: $u_{rest} = u_{reset} = 0mV$, $\tau_m = 30ms$, $u_{th} = 60mV$.

We use Spike Timing Dependent Plasticity (STDP) as a synapse learning rule to update synaptic weights. STDP is the most basic learning method in the brain, which relies on the time difference between the firing of pre-synaptic neurons and post-synaptic neurons to train the synaptic weight (Caporale and Dan, 2008). The weight update can be written as Equation 6:

$$\Delta\omega = \begin{cases} A^+ \exp(\frac{\Delta t}{\tau_+}), & \Delta t < 0 \\ -A^- \exp(\frac{-\Delta t}{\tau_-}), & \Delta t > 0 \end{cases} \quad (6)$$

Here, $\Delta t = t_i - t_j$, t_i and t_j are the time of the firing of the pre-synaptic neuron and the post-synaptic neuron, respectively (Zhao et al., 2018). A^+ and A^- are the learning rates. τ_+ and τ_- are the time constants. According to this rule, the connections will be strengthened when pre-synaptic neurons fire before post-synaptic neurons and will be weakened when pre-synaptic neurons fire after post-synaptic neurons. Here, $\tau_+ = \tau_- = 10ms$, $A^+ = 0.25$, $A^- = 0.01$.

2. The training process. We describe the training process of AE-SNN from two aspects: First, set the value of fixed weights in the three modules. In the Motor Module, the excitatory synaptic weights between SMA neurons and M1 neurons are set

so that once the SMA neurons are activated, the connected M1 neurons are activated immediately. In addition, The Emotion Module connects all neurons in the Motor Module except M1 neurons. We set these fixed synapse weights because the different emotion overt actions are determined by hereditary genetic factors (Darwin, 2015). Different emotion neurons' firing will lead to different motor neurons' firing. Since the connection between the two modules is bidirectional, these synapse weights can also be used in reverse when known the activation of the Motor Module to infer the activation of the Emotion Module.

Second, train the synapse weights between the Perception Module and the Motor Module. For the emergence of mirror neurons, we simulated the process of action execution and action re-afference and used the STDP learning rule to adjust the synapse weights. It takes 100 ms from the firing of the motor neurons to action execution, 100ms from the hearing/watching of that action to trigger the firing of the perception neurons, and the total time delay is 200 ms (Keysers and Gazzola, 2014). In the training process, the emotion neurons fire at first, the corresponding motor neurons will fire and execute the emotional overt action, and after 200 ms, the corresponding perception neurons will fire. Because action execution is a continuous process, the firing of the motor neurons is still going on, as shown in **Figure 3**. Due to this temporal correlation, the excitatory connections between the motor neurons executing the action and the perception neurons representing the same action will be strengthened by STDP, forming the mirror neurons in the Motor Module. In the emergence process of anti-mirror neurons, due to the connections between the SMA neurons and the perception neurons being inhibitory, the inhibitory connections between the SMA neurons and the perception neurons representing the same action will be strengthened by STDP. However, M1 neurons are only activated by SMA neurons, so M1 neurons are deactivated after the SMA neurons are inhibited, and the connection between the M1 neurons and the perception neurons will not be established. The M1 neurons are only activated during action execution, reproducing the property of anti-mirror neurons.

In the training stage, the connections among the three modules are established, and the mirror neurons and anti-mirror neurons emerge. When perceiving the emotional overt action of others, the perception information of overt action will be encoded in the Perception Module as input. Then it will activate the mirror neurons in the Motor Module, and further activate the corresponding emotion neurons to empathize with others.

2.4. Intrinsic Motivation for Altruistic Behavior

Waal and Preston proposed that altruistic behavior derives from intrinsic motivation, and affective empathy is a major factor to generate this motivation (Waal and Preston, 2017). Many psychologists argued that emotion is the motivation for most decisions, guiding individuals to avoid negative emotions (such as pain and sadness), which is an intrinsic motivation for survival, such as eating (Lerner et al., 2015). When people have negative emotions, they will adopt adaptive behaviors to eliminate negative emotions, as the blue arrow shown in

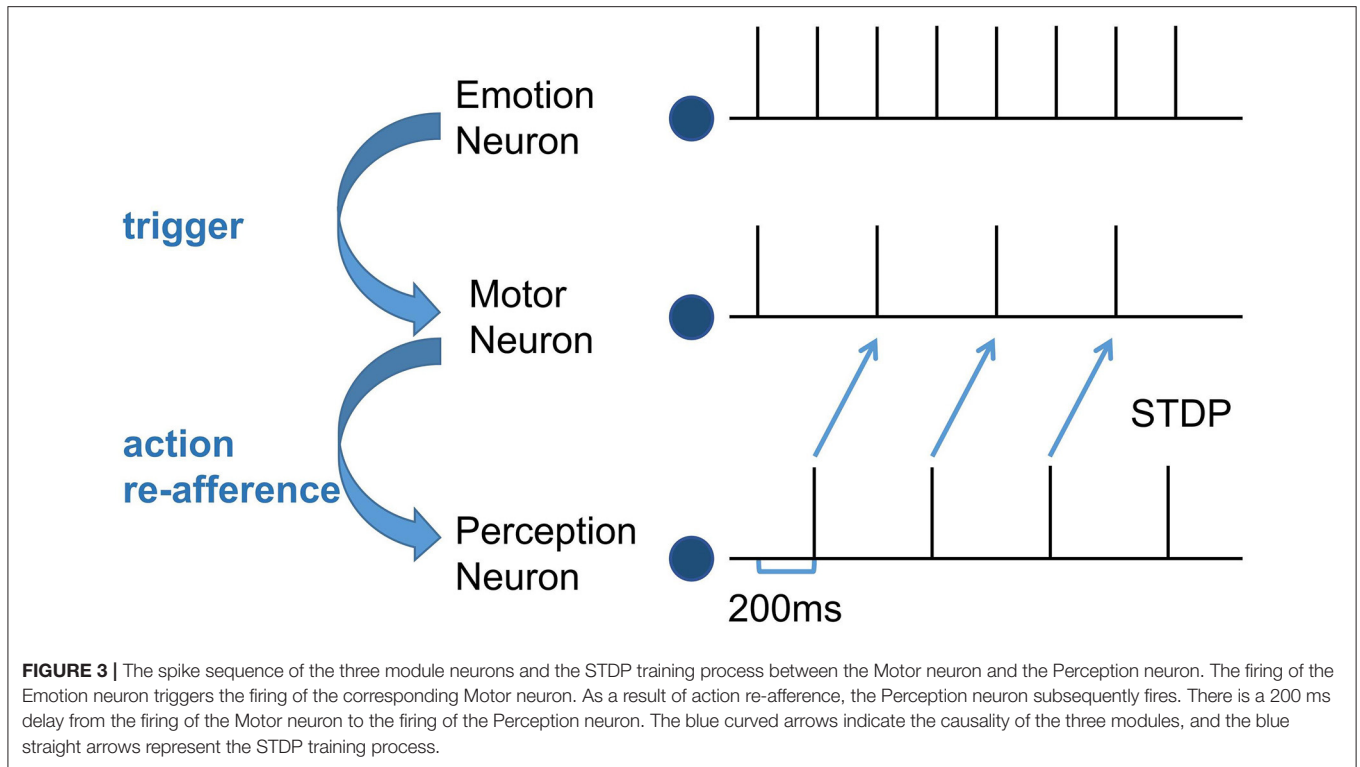


FIGURE 3 | The spike sequence of the three module neurons and the STDP training process between the Motor neuron and the Perception neuron. The firing of the Emotion neuron triggers the firing of the corresponding Motor neuron. As a result of action re-afference, the Perception neuron subsequently fires. There is a 200 ms delay from the firing of the Motor neuron to the firing of the Perception neuron. The blue straight arrows represent the STDP training process. The blue curved arrows indicate the causality of the three modules, and the blue straight arrows represent the STDP training process.

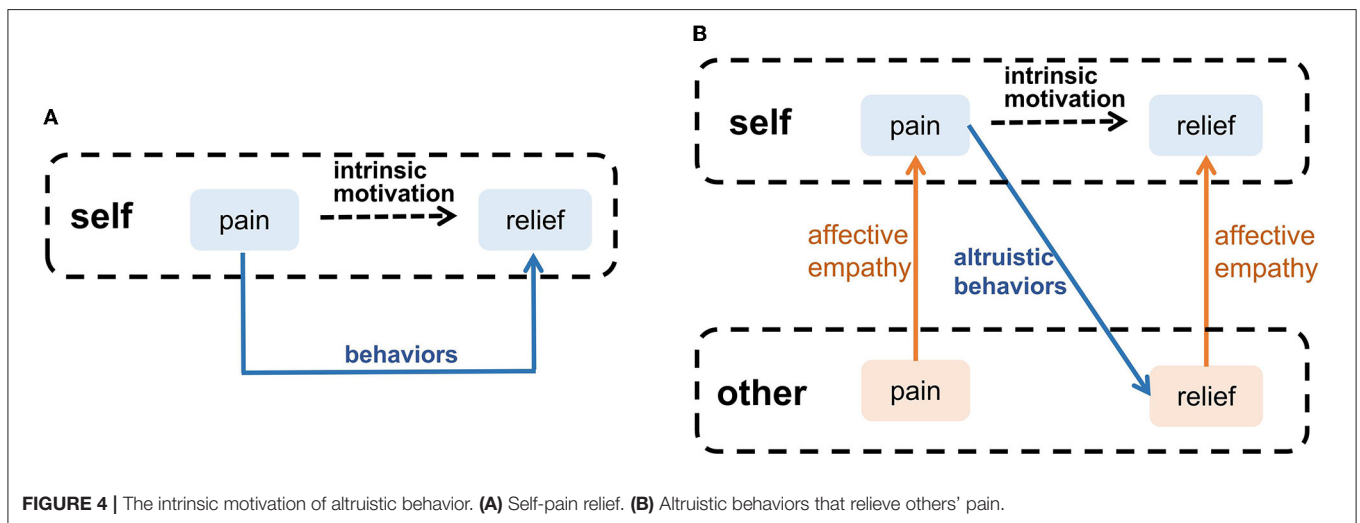
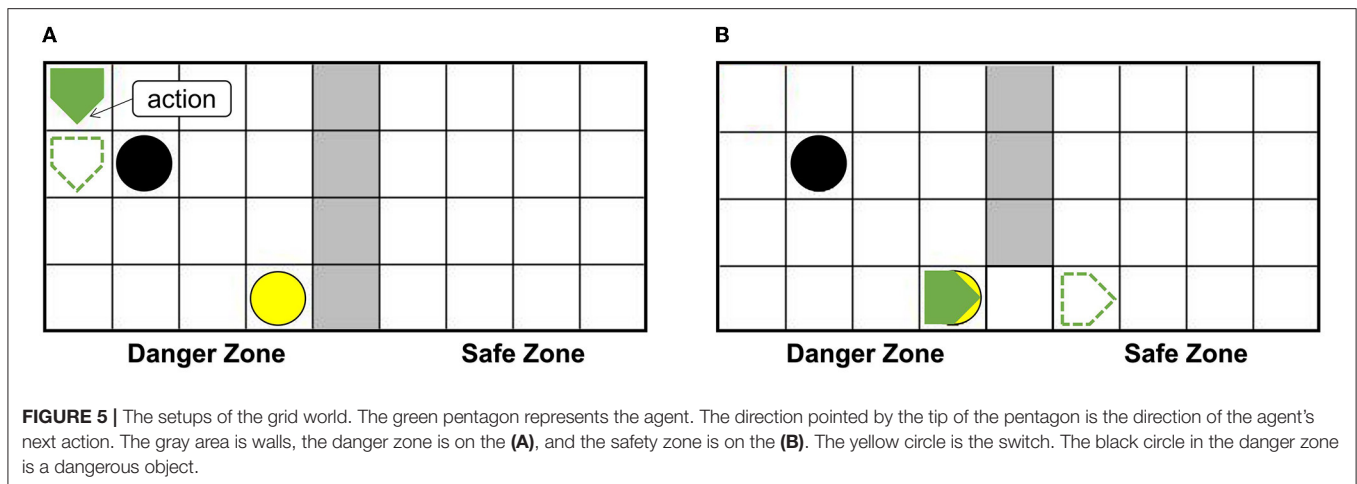


FIGURE 4 | The intrinsic motivation of altruistic behavior. (A) Self-pain relief. (B) Altruistic behaviors that relieve others' pain.

Figure 4A. When observing others' negative emotions, the ability of affective empathy can transfer the others' negative emotions to the observer, as the orange arrows shown in **Figure 4B**. This shared negative emotion provides intrinsic motivation for the observer's altruistic behavior. In order to eliminate shared negative emotion, the observer will try to adopt altruistic behavior actively, such as help, and consolation, as the blue arrow shown in **Figure 4B**. When the others' negative emotion is eliminated, the observer's shared negative emotion is

indirectly eliminated by affective empathy. Therefore, Altruistic behavior originates from spontaneous intrinsic motivation. Affective empathy can provide this motivation by realizing emotion transfer.

After exploring the intrinsic motivation of the altruistic behavior, we combine the brain-inspired affective empathy computational model proposed above to complete an altruistic rescue task and achieve the rescue of companions by an intelligent agent in the grid world. In our model, the AE-SNN



has self-other differentiation ability. When the intelligent agent has negative emotion, if the negative emotion comes from itself, it will act on itself; if the negative emotion comes from others, it will act to help the others and, thus, indirectly help itself.

3. EXPERIMENT

This subsection introduces the applications of the brain-inspired affective empathy computational model on intelligent agents in the grid world environment. Bartal et al. studied that rats learned to open doors to rescue their companions trapped in containers through trial and error (Bartal et al., 2011). We designed the experiment inspired by this behavior paradigm. The experiment process is divided into two phases: (1) The development of pain empathy ability. We designed an agentA to explore the grid world and experience pain state through the Artificial Pain Model. Then train its AE-SNN to develop its pain empathy ability. (2) The altruistic rescue task. We design a two-agent task, in which agentA uses the trained AE-SNN to empathize with agentB and implements rescue behavior driven by intrinsic altruistic motivation.

3.1. Environment Settings

The setups of the grid world are shown in **Figure 5A**. The green pentagon represents the agent. The direction pointed by the tip of the pentagon is the direction of the agent's next action, and the green dotted line represents the next position. The gray area in the middle is walls, splitting the danger zone on the left and the safety zone on the right. The black circle in the danger zone is a dangerous object, which can lead to damage when the agent crashes into it. The yellow circle is the switch. When the switch is turned on, a passage from the danger zone to the safety zone can be established as shown in **Figure 5B**, then the agent can enter the safety zone spontaneously and immediately, and the damaged agent will recover to the normal state (just like when the container is opened the trapped rat will quickly leave the container, as mentioned in the Bartal et al., 2011). In the action execution process, the agent can detect its internal state through the Artificial Pain Model. For the AE-SNN of the agent,

the number of neurons in the Emotion Module, Motor Module, and Perception Module are 40, 50, and 40, respectively. When different internal states arise, the corresponding neurons of the Emotion Module of AE-SNN will fire, activating the neurons of the Motor Module to produce corresponding emotional overt actions. In our experiments, we designed the color change of the agent to be equivalent to emotional overt actions (e.g., human facial expressions or crying). The agent shows red in a pain state and green in a normal state. Agents can use their own Perception Module to perceive their own and others' emotional overt actions, just as people can hear their own and others' cries. The trained AE-SNN makes the agent have empathy ability.

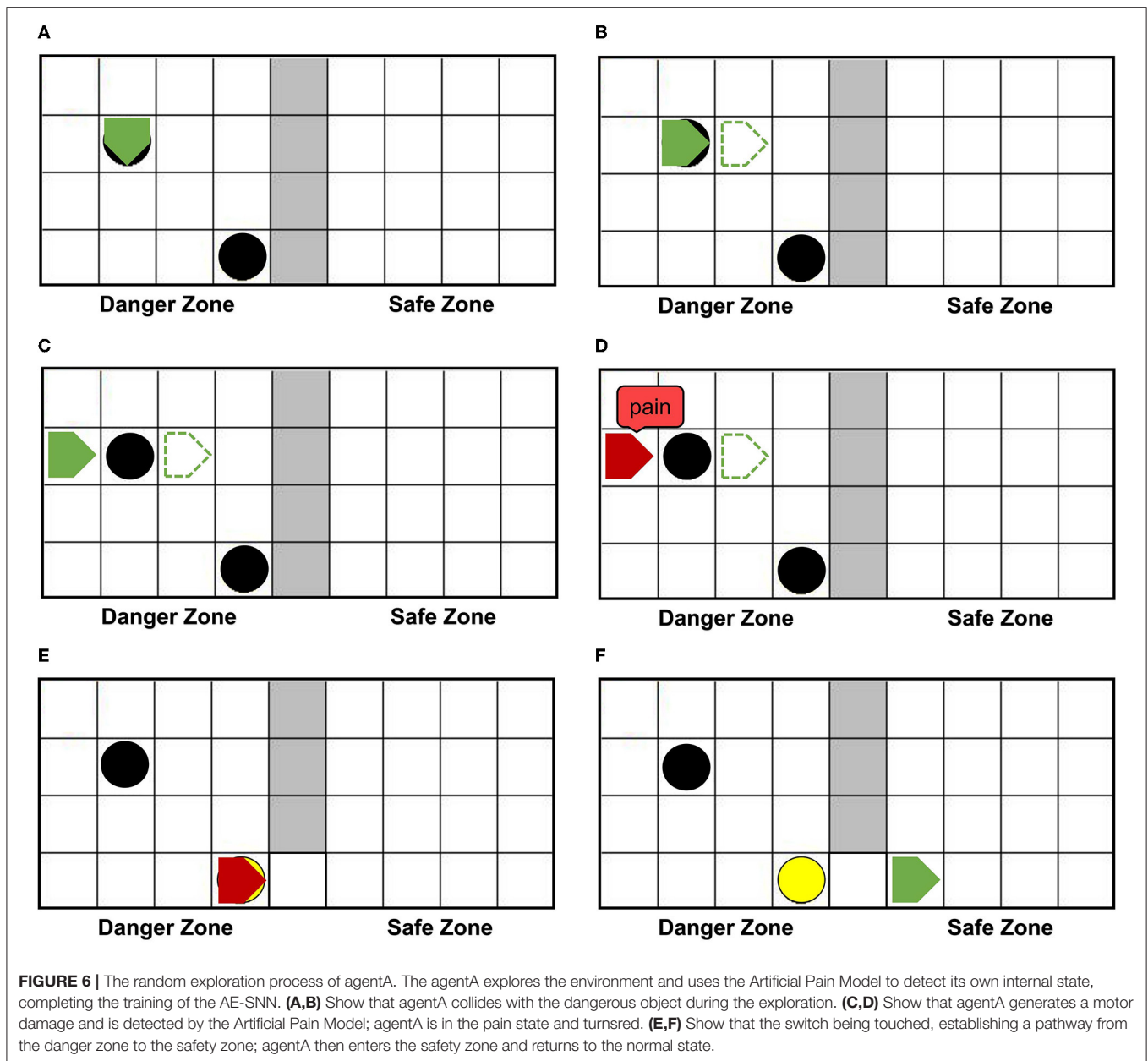
3.2. The Developing of Pain Empathy Ability

3.2.1. Experimental Procedure

We place dangerous objects, a switch, walls, and an agentA in the grid world. The agentA explores the environment and uses the Artificial Pain Model to detect its own internal state, completing the training of the AE-SNN.

We set the damage rule of agent: when an agent collides with a dangerous object, its ability to act will be impaired. For example, if a command is given to the right, the agent actually moves a unit to the left. For the Artificial Pain Model, the 'environmental dynamic' of the Free Energy will result in errors, as the third item in Equation 3. We first set the mapping relationship $g_{\phi}(\phi', a')$ of the Artificial Pain Model, which is the moving rule of the agent: When the agent performs an action, its horizontal and vertical coordinates will change accordingly. For example, if an upward action is performed, its vertical-coordinate's value reduces, and if a rightward action is performed, its horizontal-coordinate's value adds. The agent knows its moving rule in advance and uses this rule to continuously and actively predict its next position. When the prediction is wrong, the Free Energy is greater than 0, indicating that damage occurs and the agent will be in the pain state.

Figure 6A shows that agentA is colliding with the dangerous object. **Figure 6B** shows that the next motion command of agentA is toward the right, the prediction is the position of the green dotted line. But **Figure 6C** shows that the actual

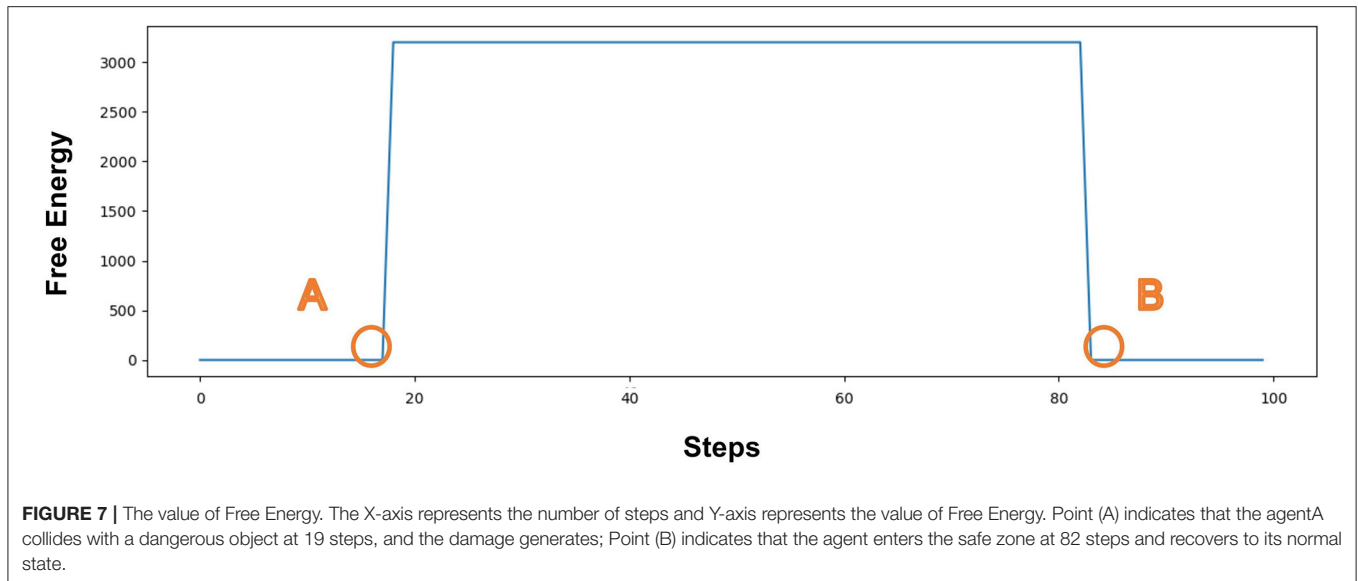


action of the agentA is toward the left. At the same time, the motor damage will be detected by the Artificial Pain Model, and the agentA will be in the pain state. After the pain state generates, the first 20 neurons in the Emotion Module that represent pain will fire and activate the neurons in the Motor Module to produce painful overt action. In this experiment, the color of the agentA turns red, as shown in **Figure 6D**. After that, the agentA's Perception Module will perceive its own color information. The first 20 neurons of the Perception Module that represent perceiving red will be activated, completing a training epoch of AE-SNN in the pain state by STDP. After 100 training epochs, the AE-SNN finishes training, and the agent has pain empathy ability. **Figures 6E,F** show that agentA touches the switch during exploration and

establishes the passage from the danger zone to the safety zone. Then it enters the safety zone and recovers to its normal state. These two figures aim to illustrate the role of the switch.

3.2.2. Results Analysis

Figure 7 represents the value of Free Energy during exploration. The X-axis represents the number of steps. The Y-axis represents the value of Free Energy. The value of Free Energy is the sum of the square of the difference between the predicted agentA's coordinates and the actual agentA's coordinates. We designed the agent as a pentagon, each vertex has a 2D coordinate (X, Y) , so the agent has five sets of 2D coordinates at each position, i.e., $(X_1, Y_1), (X_2, Y_2) \dots (X_5, Y_5)$. The Free Energy is equal to the



sum of the squares of the differences in the corresponding coordinates, as described in Equation 7. The unit grid length is 25, resulting in the Free Energy range of [3125,12500]. Since our model is only related to whether the Free Energy is greater than 0, we put all values that are greater than 3125 equal to 3125. PointA indicates that the agentA collides with a dangerous object at 19 steps, and the damage generates. The Free Energy is greater than 0, causing the agentA to be in the pain state. PointB indicates that the agentA enters the safe zone at 82 steps, and recovers its normal state. The Free Energy is equal to 0, causing the agentA's pain state to be alleviated.

$$FE = \sum_{x=1}^5 (X_{pre} - X_{actual})^2 + \sum_{y=1}^5 (Y_{pre} - Y_{actual})^2 \quad (7)$$

Figures 8A–C is the spike diagrams of the three modules of AE-SNN in the pain state. The X-axis represents the firing time, and each unit represents 100 ms. The Y-axis represents the index of the neurons. **Figure 8A** represents the firing of neurons that represent the pain state in the Emotion Module. **Figure 8B** shows the firing of neurons that represent the painful overt action in the Motor Module as the agentA's color turns red. **Figure 8C** shows the firing of neurons in the Perception Module, which corresponds to agentA perceiving its color information. The firing of the Emotion Module will then trigger the firing of the Motor Module by setting weights in advance, and then the Perception Module will fire after a 200ms delay, indicating the time interval for action execution and action re-ference. With the time correlation, the connections between the Perception Module and the Motor Module can be established by STDP.

Figure 9A shows the synaptic weights training process from the Perception Module to the Motor Module. The X-axis represents the training epochs. The Y-axis represents the value of weights. The purple line indicates the change of inhibitory weights between the Perception neurons and the SMA neurons.

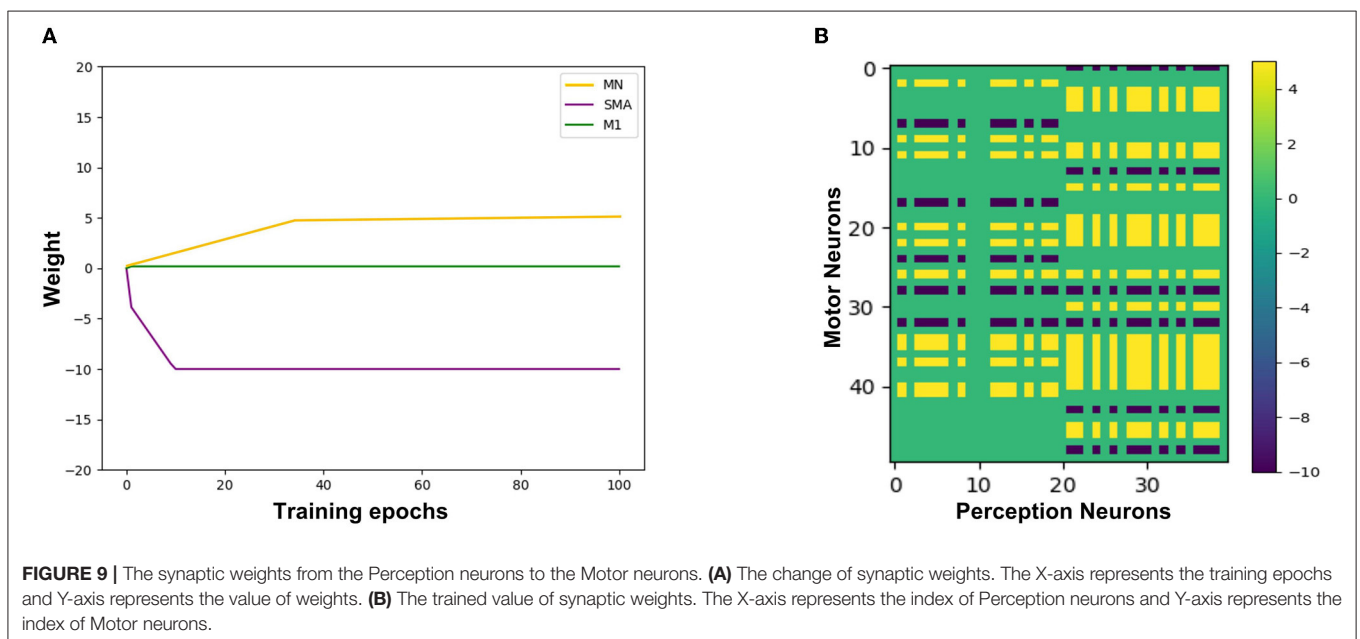
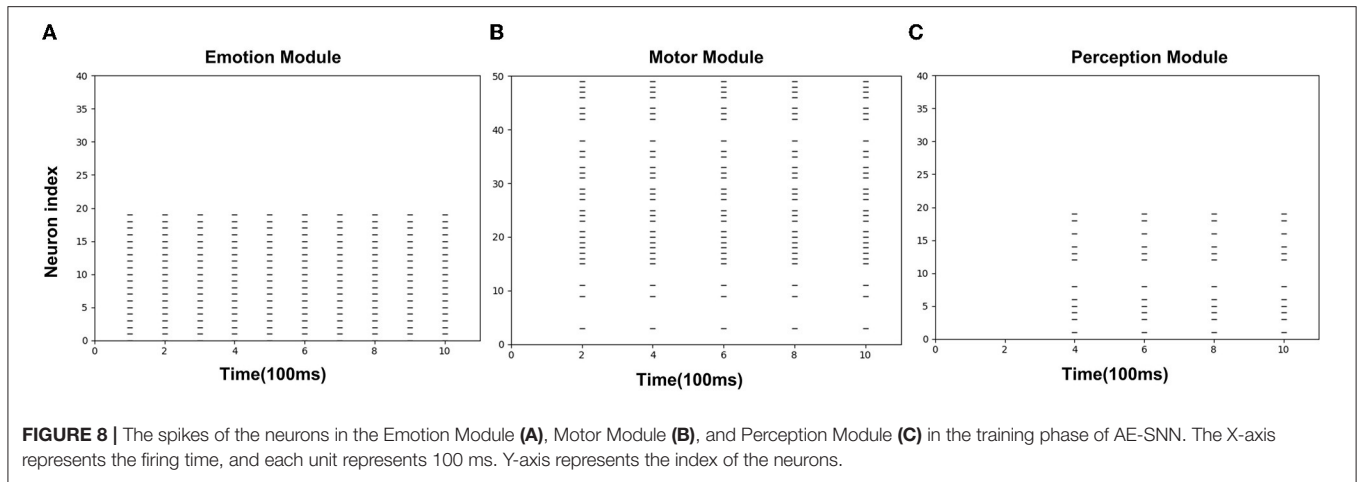
The green line indicates a change of weights between the Perception neurons and the M1 neurons, and due to the gate mechanism of the SMA neurons, the weight is maintained near 0, which is equivalent to not establishing synaptic connections. The yellow line indicates the change of excitatory weights between the Perception neurons and other neurons in the Motor Module, and the value of the weight gradually increases, indicating that these neurons are mirror neurons. **Figure 9B** shows the trained synaptic weights between the Perception Module and the Motor Module. The X-axis represents the index of Perception neurons. The Y-axis represents the index of Motor neurons. The color represents the value of the weight. The purple areas represent inhibitory synaptic weights between the Perception neurons and the SMA neurons. The yellow areas represent excitatory synaptic weights between the Perception neurons and the mirror neurons that represent the same emotional overt action. The green areas represent unestablished synaptic connections, which contain the connections between M1 neurons and Perception neurons.

3.3. The Altruistic Rescue Task

3.3.1. Experimental Procedure

Using the AE-SNN trained in Section 3.2, we designed an altruistic rescue task. **Figure 10A** shows that two agents, a dangerous object, a switch, and walls are placed in the grid world. In the previous task, the agentA with affective empathy ability is placed in the right safety zone, and agentB (green pentagon with black border) is placed in the danger zone for exploration. We place the yellow switch on the side of the safety zone, so agentB has no self-rescue ability and could only be rescued through agentA.

When agentB explores the environment, it will inevitably encounter a dangerous object, resulting in pain and turning into red color, as shown in **Figure 10B**. AgentA will perceive the color information of agentB, then generate empathy pain through AE-SNN. The intrinsic altruistic motivation will drive agentA to take



action to rescue agentB and find the optimal strategy by the reinforcement learning method. Eventually, agentA learned to go to the switch as soon as possible and establish the passage from the danger zone to the safety zone, as shown in **Figure 10C**. After the switch is turned on, agentB will enter the safety zone and relief the pain state, then agentA's empathy pain will also be relieved through AE-SNN as shown in **Figure 10D**, completing the rescue of the companion.

3.3.2. Results Analysis

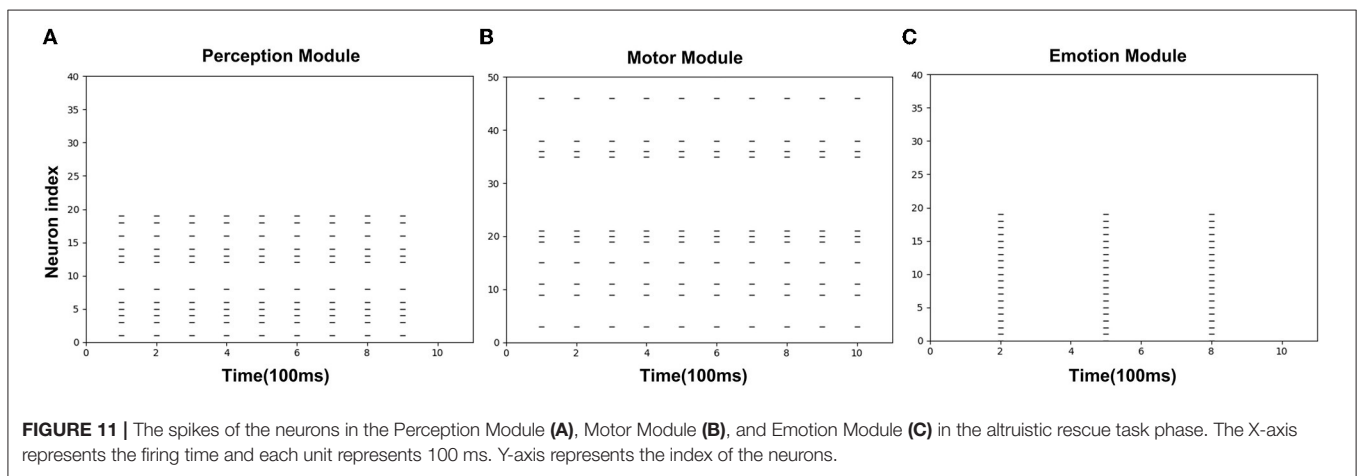
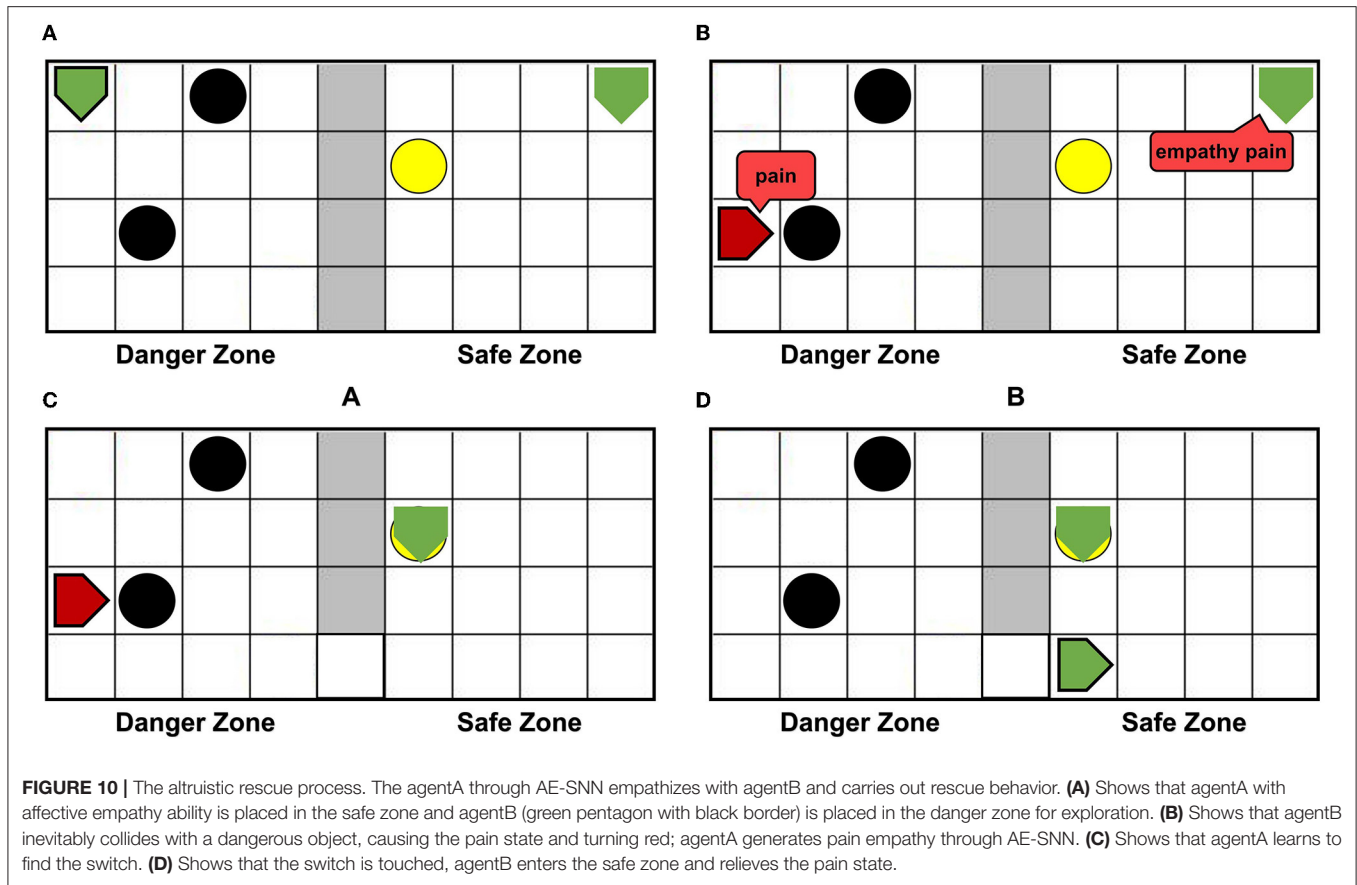
Figures 11A–C shows the spike diagrams of the three modules of the agentA during the altruistic rescue process. **Figure 11A** shows the firing of the neurons in the Perception Module, representing agentB's color information. **Figure 11B** shows the firing of the mirror neurons in the Motor Module, which realizes the understanding of the pain state at the motor level. **Figure 11C** shows the firing of the neurons in the Emotion Module, which realizes the understanding of the pain state at the emotional level.

Moreover, in this process, compared with **Figure 8B** mentioned in Section 3.2.2, not all neurons in the motor area are activated (in fact only mirror neurons are activated), so the conditions for executing painful overt actions do not arrive. Therefore, although the agentA feels pain, it does not turn red.

4. DISCUSSION

This article focuses on the brain-inspired affective empathy computational model and its application. Inspired by the neural mechanism of affective empathy, we simulated the generation process of pain, trained the emergence of mirror neurons and anti-mirror neurons, and completed the altruistic rescue task. In this section, we will discuss the strengths and limitations of our computational model and its application prospects.

In the AE-SNN, we reproduced the emergence process of mirror neurons and anti-mirror neurons. Inspired by the gating



mechanism from SMA to M1 proposed in the neuroscience literature (Christian and Valeria, 2010), we also set up neurons with similar functions in the Motor Module of AE-SNN. This eventually led to eight main types of neurons in the Motor Module, as shown in **Table 1**. If a neuron is activated when executing action1 and observing action1, it is the mirror neuron of action1, which is used to understand emotional overt action1. If a neuron is activated when executing action2 and observing action2, it is the mirror neuron of action2, which is used to understand emotional overt action2. If a neuron is activated

during the execution of both action1 and action2 and is deactivated during the observation of both action1 and action2, it is the anti-mirror neuron, which is used to distinguish between self and others and avoid unnecessary emotional overt actions in the affective empathy process.

Gazzola and Keysers studied the fMRI data of human subjects and concluded that mirror neurons were reliably observed in the premotor cortex (PM), the supplementary motor cortex (SMA), and other brain regions, and anti-mirror neurons were observed in the M1. They also discussed the gating mechanism

TABLE 1 | The analysis of the eight main neuron types in the Motor Module.

	Action1	Action2	Perception1	Perception2	Type
1	✓				
2	✓		✓		MN(1)
3		✓			
4		✓		✓	MN(2)
5	✓	✓	✓		MN(1)
6	✓	✓		✓	MN(2)
7	✓	✓	✓	✓	MN(1,2)
8	✓	✓			ANTI-MN

The MN refers to mirror neuron and the ANTI-MN refers to anti-mirror neuron. ✓ indicates whether the neuron is activated under these four conditions (Action 1, Action 2, Perception 1, Perception 2).

of the SMA neurons (Gazzola and Keysers, 2008). Mukamel et al. also illustrated a similar conclusion, and further discussed that anti-mirror neurons can maintain self-other differentiation (Mukamel et al., 2010). The results in **Table 1** show that the AE-SNN produces neuron types consistent with these two references.

In the altruistic rescue task, agentA's rescue path optimization used reinforcement learning. Unlike traditional reinforcement learning, the reward function in this task is not an extrinsic reward set by the human in advance, but an intrinsic reward obtained through the agent's internal model. First, agentB generates pain states through the Artificial Pain Model, and agentA generates empathy pain through its AE-SNN. When agentA touches the yellow switch accidentally, the passage from the danger zone to the safety zone will be established, then agentB will enter the safe zone, the pain will be relieved, and then the pain empathized by agentA will be relieved. For the brain, the relief of negative emotions can be seen as an intrinsic reward (Porreca and Navratilova, 2017). Thus, agentA marks the yellow switch as a positive reward. Then it can continuously optimize its Q-table to learn the optimal rescue path based on this reward.

There are some limitations to our study. First, there is a simplification in the design of anti-mirror neurons. In fact, the gate mechanism of SMA neurons is complex (Gazzola and Keysers, 2008), which is not easy to realize for computational modeling. Our study could be seen as the first step toward a more biologically realistic computational model of anti-mirror neurons from a functional perspective. In addition, there is some simplification in the process of artificial emotion generation. Emotion is a very complex cognitive function of organisms involving the regulation of a variety of hormones and neural circuits (Lövheim, 2012). The Artificial Pain Model proposed in this article is just a prototype. In the future, we will study more about the nature of emotion and the relationship between emotion and emotional overt actions (Mirabella, 2018; Mancini et al., 2020, 2022; Mirabella et al., 2022).

Currently, our study is only to realize the application of the mirror mechanism in a virtual environment. In the social robots field, the mirror mechanism is also necessary (Asada, 2015). If the robot associates its emotion with the perception information

of the corresponding emotion overt action, it will understand its companion's emotion when perceiving similar emotion overt information from the companion. This mirror mechanism can help robots to empathize with their companions. We will explore the application of these mechanisms in robots in the future.

The prediction process in the Artificial Pain Model simulates the evolution of pain. Organisms evolved pain by experiencing physical damage (Walters and Williams, 2019). We use the Free Energy Principle (FEP) to model this detection process of physical damage: The brain detects abnormal sensations through continuous prediction of multiple physical sensations. That is a prediction at the physical sensation level. It is worth noting that the brain also has predictions at the event level, such as predicting that the organism will be injured. This is essentially a prediction of whether an injury event will occur, which is different from the prediction at the physical sensation level in the Artificial Pain Model.

In summary, we proposed a brain-inspired affective empathy computational model, which involves the generation process of Artificial Pain and the reproduction of the mirror mechanism of MNS, along with the self-other differentiation ability. We apply this model to achieve pain empathy and altruistic behavior in the virtual grid world environment. We hope that our study will contribute to the harmonious coexistence among robot groups and between humans and robots.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HF and YZ designed the study, and performed the experiments. HF, YZ, and EL developed the algorithm and performed the result analysis, wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the New Generation of Artificial Intelligence Major Project of the Ministry of Science and Technology of the People's Republic of China (Grant No. 2020AAA0104305), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB32070100), and the Beijing Municipal Commission of Science and Technology (Grant No. Z181100001518006).

ACKNOWLEDGMENTS

We appreciate Feifei Zhao, Dongcheng Zhao, Yuxuan Zhao, and Hongjian Fang for valuable discussions. We would like to thank all the reviewers for their help in shaping and refining the manuscript.

REFERENCES

- Asada, M. (2015). Towards artificial empathy. *Int. J. Soc. Robotics* 7, 19–33. doi: 10.1007/s12369-014-0253-z
- Asada, M. (2019). Artificial pain may induce empathy, morality, and ethics in the conscious mind of robots. *Philosophies* 4, 38–30. doi: 10.3390/philosophies4030038
- Bartal, I. B.-A., Decety, J., and Mason, P. (2011). Empathy and pro-social behavior in rats. *Science* 334, 1427–1430. doi: 10.1126/science.1210789
- Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *J. Math. Psychol.* 76, 198–211. doi: 10.1016/j.jmp.2015.11.003
- Broom, D. M. (2001). Evolution of pain. *Vlaams Diergeneeskd Tijdschr.* 70, 17–21.
- Brown, E. C., and Brune, M. (2012). The role of prediction in social neuroscience. *Front. Hum. Neurosci.* 6, 147. doi: 10.3389/fnhum.2012.00147
- Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a hebbian learning rule. *Ann. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.125639
- Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502. doi: 10.1073/pnas.0935845100
- Christian, K., and Valeria, G. (2010). Social neuroscience: Mirror neurons recorded in humans. *Curr. Biol.* 20, R353–R354. doi: 10.1016/j.cub.2010.03.013
- Claret, J. A., Venture, G., and Basaez, L. (2017). Exploiting the robot kinematic redundancy for emotion conveyance to humans as a lower priority task. *Int. J. Soc. Robotics* 9, 277–292. doi: 10.1007/s12369-016-0387-2
- Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., and Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nat. Commun.* 7, 10904. doi: 10.1038/ncomms10904
- Darwin, C. (2015). *The Expression of the Emotions in Man and Animals*. Chicago, IL: University of Chicago Press.
- Davis, M. (1992). The role of the amygdala in fear and anxiety. *Ann. Rev. Neurosci.* 15, 353–375. doi: 10.1146/annurev.ne.15.030192.002033
- Donald, D. P. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science* 288, 1769–1772. doi: 10.1126/science.288.5472.1769
- Erhan, O., Mitsuo, K., and Michael, A. (2006). Mirror neurons and imitation: a computationally guided review. *Neural Netw.* 19, 254–271. doi: 10.1016/j.neunet.2006.02.002
- Erhan, O., Mitsuo, K., and Michael, A. A. (2013). Mirror neurons: functions, mechanisms and models. *Neurosci. Lett.* 540:43–55. doi: 10.1016/j.neulet.2012.10.005
- Fang, H., Zeng, Y., and Zhao, F. (2021). Brain inspired sequences production by spiking neural networks with reward-modulated stdp. *Front. Comput. Neurosci.* 15, 8. doi: 10.3389/fncom.2021.612041
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Gazzola, V., and Keysers, C. (2008). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fmri data. *Cereb. Cortex* 19, 1239–1255. doi: 10.1093/cercor/bhn181
- Ghosh-Dastidar, S., and Adeli, H. (2009). Spiking neural networks. *International Journal of Neural Systems*, 19, 295–308. PMID: 19731402. doi: 10.1142/S0129065709002002
- Hérisé, C., Khalil, R., Mofteh, M., Boraud, T., Guthrie, M., and Garenne, A. (2016). Decision making under uncertainty in a spiking neural network model of the basal ganglia. *J. Integrat. Neurosci.* 15, 515–538. doi: 10.1142/S021963521650028X
- Huang, J. Y., Lee, W. P., Chen, C. C., and Dong, B. W. (2020). Developing emotion-aware human-robot dialogues for domain-specific and goal-oriented tasks. *Robotics* 9, 31. doi: 10.3390/robotics9020031
- Jabbi, M., and Keysers, C. (2008). Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion* 8, 775–780. doi: 10.1037/a0014194
- Joffily, M., and Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Comput. Biol.* 9, e1003094. doi: 10.1371/journal.pcbi.1003094
- Keysers, C., and Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130175. doi: 10.1098/rstb.2013.0175
- Khalil, R., Karim, A. A., Khedr, E., Mofteh, M., and Moustafa, A. A. (2018a). Dynamic communications between gabaa switch, local connectivity, and synapses during cortical development: a computational study. *Front. Cell Neurosci.* 12, 468. doi: 10.3389/fncel.2018.00468
- Khalil, R., Mofteh, M. Z., and Moustafa, A. A. (2017). The effects of dynamical synapses on firing rate activity: a spiking neural network model. *Eur. J. Neurosci.* 46, 2445–2470. doi: 10.1111/ejn.13712
- Khalil, R., and Moustafa, A. A. (2022). A neurocomputational model of creative processes. *Neurosci Biobehav Rev.* 137:104656. doi: 10.1016/j.neubiorev.2022.104656
- Khalil, R., Tindle, R., Boraud, T., Moustafa, A. A., and Karim, A. A. (2018b). Social decision making in autism: on the impact of mirror neurons, motor control, and imitative behaviors. *CNS Neurosci. Therapeut.* 24, 669–676. doi: 10.1111/cns.13001
- Kilner, J. M., and Friston, K., and Frith, C. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Process* 8, 159–166. doi: 10.1007/s10339-007-0170-2
- Lamm, C., Decety, J., and Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* 54, 2492–2502. doi: 10.1016/j.neuroimage.2010.10.014
- Lanillos, P., and Cheng, G. (2018). “Adaptive robot body learning and estimation through predictive coding,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 4083–4090.
- Lee, H. S., and Kang, B. Y. (2020). Continuous emotion estimation of facial expressions on jaffe and ck+ datasets for human-robot interaction. *Intell. Serv. Robotics* 13, 15–27. doi: 10.1007/s11370-019-00301-x
- Leite, I., Castellano, G., and Pereira, A. (2014). Empathic robots for long-term interaction. *Int. J. of Soc. Robotics* 6, 329–341. doi: 10.1007/s12369-014-0227-1
- Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015). Emotion and decision making. *Ann. Rev. Psychol.* 66, 799–823. doi: 10.1146/annurev-psych-010213-115043
- Lou, A. (2011). *Empathy and Sympathy in Ethics*. Internet Encyclopedia of Philosophy. Available online at: <http://www.iep.utm.edu/empsymp/> (accessed July 1, 2022).
- Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Med. Hypotheses* 78, 341–348. doi: 10.1016/j.mehy.2011.11.016
- Malin, A., Jakob, E., and Eric, M. H. (2011). A comparison of empathy for humans and empathy for animals. *Anthrozoös* 24, 369–377. doi: 10.2752/175303711X13159027359764
- Mancini, C., Falciani, L., Maioli, C., and Mirabella, G. (2020). Threatening facial expressions impact goal-directed actions only if task-relevant. *Brain Sci.* 10, 794. doi: 10.3390/brainsci10110794
- Mancini, C., Falciani, L., Maioli, C., and Mirabella, G. (2022). Happy facial expressions impair inhibitory control with respect to fearful facial expressions but only when task-relevant. *Emotion* 22, 142–152. doi: 10.1037/emo0001058
- McGreyor, S., Baltieri, M., and Buckley, C. L. (2015). A minimal active interference agent. *arXiv [Preprint]*. arXiv: 1503.04187. doi: 10.48550/ARXIV.1503.04187
- Melzack, R. (2001). Pain and the neuromatrix in the brain. *J. Dent. Educ.* 65, 1378–1382. doi: 10.1002/j.0022-0337.2001.65.12.tb03497.x
- Mirabella, G. (2018). The weight of emotions in decision-making: How fearful and happy facial stimuli modulate action readiness of goal-directed actions. *Front. Psychol.* 9, 1334. doi: 10.3389/fpsyg.2018.01334
- Mirabella, G., Grassi, M., Mezzarobba, S., and Bernardis, P. (2022). Angry and happy expressions affect forward gait initiation only when task relevant. *Emotion*. doi: 10.1037/emo0001112
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., and Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13, 684–701. doi: 10.1006/nimg.2000.0715
- Mukamel, R., Ekstrom, A., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Curr. Biol.* 20, 750–756. doi: 10.1016/j.cub.2010.02.045

- Porreca, F., and Navratilova, E. (2017). Reward, motivation and emotion of pain and its relief. *Pain* 158(Suppl. 1), S43. doi: 10.1097/j.pain.0000000000000798
- Preston, S., and de Waal, F. (2017). Only the pam explains the personalized nature of empathy. *Nat. Rev. Neurosci.* 18, 769. doi: 10.1038/nrn.2017.140
- Ramstead, M. J. D., Badcock, P. B., and Friston, K. J. (2018). Answering schrödinger's question: a free-energy formulation. *Phys. Life Rev.* 24, 1–16. doi: 10.1016/j.plrev.2017.09.001
- Reep, R. L., Finlay, B. L., and Darlington, R. B. (2007). The limbic system in mammalian brain evolution. *Brain Behav. Evol.* 70, 57–70. doi: 10.1159/000101491
- Rizzolatti, G., and Luppino, G. (2001). The cortical motor system. *Neuron* 31, 889–901. doi: 10.1016/S0896-6273(01)00423-8
- Rizzolatti, G., and Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nat. Rev. Neurosci.* 17, 757–765. doi: 10.1038/nrn.2016.135
- Shamay-Tsoory, S. G., Judith, A. P., and Daniella, P. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132(Pt 3), 617–627. doi: 10.1093/brain/awn279
- Soujanya, P., Erik, C., Rajiv, B., and Amir, H. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Inf. Fusion* 37, 98–125. doi: 10.1016/j.inffus.2017.02.003
- Tal, D., and Schwartz, E. L. (1997). Computing with the leaky integrate-and-fire neuron: Logarithmic computation and multiplication. *Neural Comput.* 9, 305–318. doi: 10.1162/neco.1997.9.2.305
- Waal, F., and Preston, S. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nat. Rev. Neurosci.* 18, 498–509. doi: 10.1038/nrn.2017.72
- Walters, E. T., and Williams, A. C. d. C. (2019). Evolution of mechanisms and behaviour important for pain. *Philos. Trans. R. Soc. B Biol. Sci.* 374, 20190275. doi: 10.1098/rstb.2019.0275
- Watanabe, A., Ogino, M., and Asada, M. (2007). Mapping facial expression to internal states based on intuitive parenting. *J. Rob. Mechatron.* 19, 315–323. doi: 10.20965/jrm.2007.p0315
- Wiech, K., and Tracey, I. (2013). Pain, decisions, and actions: a motivational perspective. *Front. Neurosci.* 7, 46. doi: 10.3389/fnins.2013.00046
- Woo, J., Botzheim, J., and Kubota, N. (2017). Emotional empathy model for robot partners using recurrent spiking neural network model with hebbian-lms learning. *Malaysian J. Comput. Sci.* 30, 258–285. doi: 10.22452/mjcs.vol30no4.1
- Yamada, H. (1993). Visual information for categorizing facial expression of emotions. *Appl. Cogn. Psychol.* 7, 257–270. doi: 10.1002/acp.2350070309
- Zhao, F., Zeng, Y., and Xu, B. (2018). A brain-inspired decision-making spiking neural network and its application in unmanned aerial vehicle. *Front. Neurobot.* 12, 56. doi: 10.3389/fnbot.2018.00056
- Zheng, M., She, Y., Liu, F., Chen, J., Shu, Y., and XiaHou, J. (2019). “Babebay-a companion robot for children based on multimodal affective computing,” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu: IEEE), 604–605.
- Zipser, K., Lamme, V. A. F., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389. doi: 10.1523/JNEUROSCI.16-22-07376.1996

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Feng, Zeng and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.