BENTHAM OPEN

CrossMark

The Open Cybernetics & Systemics Journal

RESEARCH ARTICLE

EMOTION-III Model: A Theoretical Framework for Social Empathic Emotions in Autonomous Control Systems

Nicoladie D. Tam[*]

Department of Biological Sciences, University of North Texas, Denton, Texas 76203, USA

Abstract: A theoretical model of social empathic emotion is derived based on the principles of survival by extending the maximization of self-gains to include others as an extended-self. This extended-self model of optimization for survival provides the computational mechanisms in the optimization process to maximize self-gains without minimizing the gains (or maximizing losses) for the other individuals. Thus, it can resolve conflicts in a competitive environment, and change the social dynamics into a cooperative interaction instead. The social emotion is emerged as a feedback mechanism for self-detection and self-correction of the disparity between self and others in the optimization process for self-preservation and survival. Maximization of the desirable gains will increase the survivability. Social emotion, such as empathy, emerges as the emotional feedback for the optimization process in survival by extending a self-centered frame of reference to an inclusive extended-self frame of reference. By including other individuals as the extended-self in the optimization process for self-preservation, it reduces the conflict of maximizing losses for the other person. Four types of social interaction in optimization strategy for survival are also discussed in relation to cooperation, competition, commensalism and altruism. This provides the theoretical foundation for the EMOTION-III model in deriving social empathic emotion that incorporates other individuals in the optimization process. This extends the previous two models of EMOTION-I and EMOTION-II, which only deals with self-emotion (*i.e.*, emotions based on self-survival and self-preservation without taking other individuals or social interactions into account).

Keywords: Autonomous control system, Emotional model, Empathy, Neural processing, Optimization, Social emotions.

1. INTRODUCTION

Toward the understanding of how social dynamics and emotions are related to each other in social interactions, a computational model of social emotion is developed to account for the social dynamics of cooperation and competition. The computational model is derived based on the principles of optimization to maximize/minimize the gains/losses in these interactions within a social group.

A "social group" is a group of individuals who interact with each other to form a system that depends on each other for survival. A social group often requires interaction of its members to function as a unit toward a common desirable goal [1]. For instance, a social group (such as a couple or a family) often operates as a system rather than as an individual in its survival as a whole. Similarly, a social group in animals operates as a pack in hunting toward a common goal for survival. There are many different types of interaction that will produce different types of behavior in survival, such as cooperative and competitive behavior.

Although autonomous robots have incorporated emotion in the control of autonomous behavior [2 - 6], most of these autonomous robots *mimic* the human emotions as an "add-on" feature to the unemotional cognition rather than incorporate emotions as an integral part of the self-adaptive system in the feedback control for social behaviors. Although there is a parallel between the human brain and the artificial emotional system in robots [7], such comparison

---
* Address correspondence to this author at Department of Biological Sciences, University of North Texas, Denton, Texas 76203, USA; Tel: 940-565-3261; Email: Nicoladie.Tam@unt.edu

does not address the computational mechanisms of emotions in self-adaptive autonomous control, or in cooperative/competitive behavior in robots.

In order to understand the dynamics of interaction, it is important to reveal the underlying mechanisms involved in producing such cooperative or competitive behaviors. This paper will examine and derive the principles underlying different social interactions using social emotion as the emergent property that regulates such behavioral dynamics. The emergence of social emotion in social interaction serves as the computational mechanisms for producing either cooperative or competitive interactions.

An example of social emotion is empathy, in which empathic emotion can only exist in the presence of another member in the social group. Empathy cannot exist in isolation in absence of other members. The evolution of empathic emotion provides the common thread in which the social interaction can allow the individuals to function as a functional unit (*i.e.*, as a system rather than in isolation) [8].

The traditional psychological view of empathy is the ability to feel for another person by emotional attainment [9]. On the other hand, we seek to derive the operating principles of social interactions that result in empathy by applying the computational principles in autonomous behaviors interacting in a cooperative social group for survival. In doing so, we will create a model in which such emotional attainment is accomplished by applying the neuroengineering principles for creating an interacting autonomous system working towards a common goal among its members of the social group.

2. COMPUTATIONAL PRINCIPLES FOR DIFFERENT TYPES OF SOCIAL INTERACTION IN COOPERATION, COMPETITION AND COMMENSALISM

This paper will derive the computational mechanisms for the emergent of social emotions in the process of optimization for gains/losses in these social interactions for survival. The type of interactions can be either co-dependent, inter-dependent or independent of each other, which will result in different type of social dynamics, such as cooperation, competition and commensalism. The survival of a group (as well as its members) is dependent on the dynamics of these interactions, which will be explored in this paper.

2.1. Principles of Group Dynamics

Similarly, the control of autonomous robots also evolves from the control of an individual robot to a group of swarm robots that interact and function together to form a social group. Although each of these robots can operate autonomously and independently, the interaction among the swarm robots can produce behavior that no single robot can accomplish without the other. Thus, the function of an autonomous robot can extend beyond itself to include other robots in its operation, even though each robot operates independently.

2.2. Principles of Extending Self to Include Others

Similarly, the behavior of a human individual can extend beyond itself to include other individuals (members of a social group) in its functioning. Generalizing these principles of interaction, the theory of social evolution can be applied not only to humans, but also other animals and autonomous robots computationally, without loss of generality. Thus, in general, each of these autonomous individuals can be considered as an agent in the social interaction. The survival function of one agent can affect the survival of another agent, when they interact with one another.

2.3. Computational Principles for Optimization in Survival

Specifically, the computational problem to be solved in social interaction is that it requires optimizing the competing solutions among members for survival. When the survival of an individual depends on the survival of other members of the group, then the optimization strategy among multiple members in a social group is also required. There are many different types of interaction that vary from cooperation, to competition and commissural, depending on the dynamics of the interaction. These dynamics of interactions can be beneficial, detrimental or neutral to the group (as a whole) and/or the members (individually).

2.4. Computational Principles in Creating Different Types of Social Interactions

In order to explore the dynamics of such a socially interactive group, the computational problem can be posed as an optimization problem of gains and losses relative to the individuals, as well as relative to the group. The optimization problem is simply maximizing the gains and minimizing the losses.

2.5. Principles of Optimization Strategies in Resulting in Different Social Interactions

The types of social interactions can be subdivided according to the optimization strategies:

- Cooperative social interaction,
- Competitive social interaction,
- Commensalistic social interaction, and
- Altruistic social interaction.

2.5.1. Cooperative Social Interaction

The optimization is to maximize the gains for both an individual and its group members. The results of such interactions often produce benefits to all members. Note that a cooperative interaction includes collaboration and mutualism.

2.5.2. Competitive Social Interaction

The optimization is maximizing the gains for an individual, which also minimizes the losses for the same individual. At the same time, it may also require the minimization of the gains for the other group members, which could result in maximizing the losses for them. The results of such interactions often produce benefits to the individual, but detriment to the other members. Note that a competitive interaction includes predation, parasitism in animals and war in human.

2.5.3. Commensalistic Social Interaction

The optimization is maximizing the gains for an individual without affecting the other group members. The results of such interactions often produce benefits to the individual, but do not affect the other members. This usually means that there is little or no interaction with the unaffected members.

2.5.4. Altruistic Social Interaction

The optimization is minimizing the gains for an individual while maximizing the gains for others. The results of such interactions often result in detrimental to the individual, but beneficial the other members. This usually means self-sacrifice, in which the individual could die for others in order to save other's life in altruism.

The four types of social interaction can be summarized in Table **1** showing the differences in the outcome of these social interactions.

Table 1. Outcomes of four different social interactions depending on whether the interactions are cooperation, competition, commensalism or altruism.

Interactions	Self	Others
Cooperation	beneficial (+)	beneficial (+)
Competition	beneficial (+)	detrimental (–)
Commensalism	beneficial (+)	unaffected (0)
Altruism	detrimental (–)	beneficial (+)

The outcomes of the social interaction are dependent on the optimization principles in survival in relation to self and others:

- If the survival of self and others is maximized in the optimization process, then this leads to cooperative behavior in which the outcome is beneficial to both self and others (see Tables **1-3**).
- If the survival of self is maximized and the survival of the others is minimized, then this leads to competitive behavior in which the outcome is beneficial to self but detrimental to others (see Tables **1-3**).
- If the survival of self is maximized, while the survival of the others is not considered, then this leads to commensalism behavior in which the outcome is beneficial to self while the others is unaffected (see Tables **1-3**).
- If the survival of self is minimized, while the survival of the others is maximized, then this leads to altruistic behavior in which the outcome is detrimental to self, while beneficial to the others. This is a form of reverse competition (see Tables **1-3**).

Table 2. Optimization principles for survival in four different social interactions depending on whether the interactions are cooperation, competition, commensalism or altruism.

Interactions	Self	Others
Cooperation	maximization (+)	maximization (+)
Competition	maximization (+)	minimization (−)
Commensalism	maximization (+)	none (0)
Altruism	minimization (−)	maximization (+)

Finally, since the optimization strategy employed in the social interaction is either maximization or minimization, these four types of interaction can lead to either gain or loss as the outcome.

Table 3. Outcomes of optimization principles for survival in four different social interactions depending on whether the interactions are cooperation, competition, commensalism or altruism.

Interactions	Self	Others
Cooperation	gain (+)	gain (+)
Competition	gain (+)	loss (−)
Competition	gain (+)	neutral (0)
Commensalism	loss (−)	gain (+)

3. THE "EXTENDED-SELF" EMPATHIC MODEL

A computational social interaction model is developed to incorporate the empathic social emotion as an implicit optimization variable to extend the concept of "*self*" to include "*others*" (as a part of the "*extended-self*"). This simple extension of *self* that includes *others* can be used to achieve any type of social interaction, not just the cooperative and commensal interactions, but also the competitive interaction.

This "extended-self model" uses the optimization principles in survival to maximize the desirable gains and minimize losses for either self or others. In the *self-centered model* of survival, the optimization of desirable gains is limited to the *self*-gains, without regards to the gains or losses of the *others*. This principle of optimization can function if the individual is operating in isolation, without interacting with others. But in most social environments, where the action of one member can often affect the others, then the optimization process for desirable gains may require inclusion of the other members in its computation.

In this *extended-self model*, it includes other members of the social group as a part of the *self* in its optimization process for *self*-gains. In other words, it includes other members of the social group in the process of maximizing the desirable gains, so that the survival of the social group as a whole will increase (rather than increasing the survival of oneself only). Even though each of the individuals operates independently, when the optimization strategy for survival includes the survival of the other group members or the entire group (as a system), then the self-centered model is expanded into the extended-self model in the computational optimization strategy.

3.1. Extended-self Model is not the Same as Other-centered Model in Altruism

Note that the extended-self model is different from the other-centered model, in which the optimization is focused on the survival of the other member at the expense of oneself. This leads to altruistic behavior.

Altruism is defined as conferring benefits to others at the expense of self [10]. This often leads to detriment to self while benefiting others in the social group, when the optimization for survival is other-centered rather than self-centered.

On the other hand, the extended-self model includes both self and other members in the optimization for survival, such that both parties in the social unit will survival as a whole. This extended-self model is different from the other-centered model with respect to the frame of reference in the optimization process.

3.2. Neurobiological Basis of Empathic and Altruistic Behavior

Numerous studies of empathy in humans have linked the hormone oxytocin with the empathic behavior [11 - 13] that involves the insular cortex and the anterior cingulate cortex for processing [14]. It also involves the sensori-motor component for empathic pain [15]. Yet these studies only address the empathic phenomena in the human brain and behavior without addressing the computational role of empathy for survival of not just self, but also other individuals.

4. OPTIMIZATION STRATEGIES FOR SELF-PRESERVATION

The principle of self-survival is based on the self-preservation principle for maximizing the desirable gains for the self to order to increase its survivability. If an individual is functioning independently on its own, without affecting others, then the optimization is a simple process for maximizing the desirable gains of that individual only. But if the maximization process affects some other individuals, then there are many different strategies in which the optimization can be done, depending on the desirable outcomes.

4.1. Depth-first Search *vs*. Breadth-first Search Principles

If the objective is to find an optimal solution for the survival of the social group as a whole (as opposed to the survival of the individuals), then there are many optimization strategies to achieve such goals. The optimal solution can be explored by either a depth-first search or a breath-first search strategy among the members, which is analogous to the search strategy in game theory. If the survival of the entire group is essential, then the optimization strategy is to maximize the desirable gains of the entire group collectively, rather the gains of the members individually.

4.2. Optimization Strategies for the Survival of a Social Group

When the maximization of self-gain also affects the gain/loss of other members, then the principle of self-preservation requires that the optimization for the gains for oneself take precedence over the gains for others. However, when the goal of the optimization is to maximize gains for *self* over *others*, it will result in the competitive social interaction. This maximization of gains for *oneself* becomes selfishness, while the maximization of losses for *others* becomes combative in social interactions.

Based on this optimization principle, cooperative social interactions cannot be achieved when *self*-regarding interests take priority over *other*-regarding interests. The consequences of this combative interaction often lead to destruction of other group members in social competition rather than mutual preservation in social cooperation.

4.3. Evolution of the Emergent Empathic Emotion

In order to achieve cooperative behavior (while not violating the optimization principle for self-preservation), the *empathic emotion* is derived in this model computationally as a "social emotion" in social interactions where the empathic emotion exists only when there is an inclusion of others. In other words, empathy can only exist by considering others in the emotional feedback; empathy does not exist without the consideration of others. Thus, empathy is an emotion in which the individual extends the boundary of *self* to include other members as a part of the *extended-self* in its optimization strategy [16, 17].

When other members are included as a part of the extended-self, then the optimization strategy can be done to maximize gains for both self and others simultaneously, without compromising the survival of the other members or the social group as a whole. An example of such an extended-self model in a social group is a couple in a marriage, which is formed as a unified group to include the partner as a part of the extended-self. This inclusive principle allows the self-preservation principle to maximize gains for not just own self, but for the extended-self (without excluding the other members of the social group). Using the example above, a married couple is a social group that is formed in order to maximize the gains for *both* partners by extending oneself to include the partner instead of maximizing the gains for oneself only while ignoring the gains of the partner. Thus, this prevents maximizing losses for others (because the extended-self now includes both self and others implicitly).

This extended-self model provides the basis for the evolution of empathy and empathic emotions, which is the ability to feel for others. This extension is done computationally by incorporating others as a part of the original self to achieve empathy.

4.4. Derivation of the Empathic Principle

Empathy is defined as the ability to feel for others. By extending oneself to include others as a part of the extended-self, the ability to feel for others is essentially the ability to feel for oneself, except that the newly formed self includes others as a part of the extended-self.

This empathic emotion also serves as the basis for compassion [18], which is a *motivated* emotion that not only feels for others, but also motivates actions to minimize the losses for others. This resolves the dilemma of maximizing the losses for others in the process of maximizing gain for self in the self-centered emotional model.

4.5. Derivation of the Altruistic Principle

On the other hand, the other-centered model (the "self-sacrifice model") often maximizes the gains for others at the expense of self, which leads to altruistic behavior (*i.e.*, self-sacrificing behavior). Thus, self-sacrifice is an optimization process that occurs when the maximization of the gains for others take precedence over the gains for self. This leads to altruism, which increases the survival of others at the expense of self. Thus, the outcome is detrimental to self, while beneficial to others. This is a form of reverse competition principle.

5. THE EMOTION-III MODEL

The social emotion model (the EMOTION-III model) is an extension of the computational models of EMOTION-I [19] and EMOTION-II [20], which describe the evolution of emotions from self-centered emotions to empathy based on the basic survival principles. The survival consideration extends now from a self-centered model to an extended-self model (to include others) by providing emotional feedback that is computed relative to the survival of the extended-self (which includes others as a result).

The computational theories of optimization for gains/losses in emotional responses had been validated experimentally in human subjects with respect to the different emotions, including; happiness [21, 22], anger [23], sadness [24], and jealousy [25], as well as fairness perception [26 - 28]. It has been shown that the intensity of emotion is directly proportional to the gains/losses relative to the self-centered model.

The present social emotion model extends the previous self-centered models to include other individuals as a part of the extended-self in the derivation of social emotions. It extends the model by a simple extension of the optimization variable to include others as a part of the self in the mathematical derivations of self-emotions described earlier [16, 17, 29]. These self- and social-emotion models provide the computational principles for resolving the optimization problems in maximizing gains for oneself, without necessarily maximizing losses for others that often occurs in a competitive social environment. This resolves the dilemma that often creates the conflict between oneself and others, if the maximization of gains is considered without the regards for others.

This extended-self model of empathy and compassion can now be used to explain the social behaviors in maternal love and romantic love using a computational optimization model. This can be done without requiring any other psychological principles or anthropological rationales for the evolution of empathy, love and cooperative behaviors in social interactions. Maternal love is empathic emotion resulted in the mother-child interaction as a social group of parent and child (offspring) by extending oneself to include the child. Romantic love is the empathic emotion resulted in the social interaction of "pair bonding" in a social group of romantic partners by extending oneself to include the romantic partner.

Emotion is one of the psychological phenomena, which is often considered as subjective and difficult to quantify. Yet, it plays an important role in affecting our behavior and social interactions. Since emotion is often a subjective measure, it is essential to quantify emotion based on a mathematical model so that the computational principles involved in emotional process can be understood. Most importantly, quantification of the emotional biases using objective measures will allow us to determine how emotion alters other computational processes, such as the decision-making process or fairness judgment [22, 26, 27, 29 - 31].

5.1. Computational Models of Emotion

In deriving a computational model of emotion, we choose a minimalistic model (using minimal assumptions) to account for the phenomena that can be observed behaviorally, in order to avoid making any preconceived notion about emotion that may bias the objective derivation. The computational model will provide testable hypotheses to validate against these behavioral observables experimentally. The next goal after developing such computational models is to verify whether such model explains the neuropsychological phenomena - whether neural correlates correspond to the hidden variables affecting the emotional processing. If such correlates were found, then the next step is to quantify these hidden variables as an objective measure for detecting how emotional biases are skewed with respect to these hidden variables. Experimental evidence in human subjects has been validated the emotional theory in terms of the self-emotions (*i.e.*, emotions generated based on self-regarding concerns without any other-regarding concerns), including include happy [21, 22], angry [23], sad [24], and jealous emotion [25]. Thus, the theoretical derivation of an computational model of social emotion will also provide an objective model for designing experiments to test these hypotheses in humans in the subsequent studies.

5.2. Extension of the EMOTION-I and EMOTION-II Models

The present empathic emotion model incorporates self-regarding and other-regarding concerns in the optimization process by extending the boundary of the *self* from an individual entity to include other members as the extended-self. It is the progressive refinement of the original EMOTION models to provide a step-by-step evolution of the emotion - from the emotional "feel" of how pleasant the sensation is as desirable emotional valence in EMOTION-I Model [19] - to the self-discovered feedback in error detection between desirable gains and losses in EMOTION-II Model [20]. These models are relying on the self-feedback as an internal guide to correct disparity errors autonomously. The EMOTION-III Model will incorporate social feedback as a guide for self-discovering disparity errors between the individuals of the social group. The naming of these models is abbreviated by the acronym EMOTION, which stands for the "Emotional Model of the Theoretical Interpretations of Neuroprocessing."

In deriving the theoretical basis for the evolution of social emotion, social emotion is an extension of the self-emotion that includes not only self, but also others as a part of oneself in processing emotions. Emotion is in this model is evolved as a self-discovered error-correction feedback process to assess the accuracy of the predictions generated internally compared to the reality. The probability of survival of an organism is dependent on the accuracy of the predictions. The more accurate the prediction is, the higher the chance of survival. Therefore, an autonomous organism has to derive a measure to assess how accurate the prediction is. Emotion in this model is derived from the disparity between the predicted outcomes and the actual outcomes. Therefore, this discrepancy measure can be used to correct the inaccuracy of the internal model for correction. Emotion is essentially a measure of this discrepancy between the prediction and the actuality.

5.3. Brief Review of EMOTION-I Model for Identifying the Emotional Feel of Sensation

In brief, the EMOTION-I [19] model derived the pre-processing computations for the contextual "feel" of sensory stimuli for survival, so that the "pleasantness" of the sensory input (sensation), which does not simply encode the intensity (quantity) of the stimulus, but also the quality of the stimulus, *i.e.*, whether it is pleasant or unpleasant. That is, it encodes how pleasant the sensation is so that it can be used as a feedback to increase the survivability by avoiding potential danger or damages if the sensory stimuli were unpleasant. This contextual feel of the sensation is the first step in the pre-processing of the emotional inputs by abstracting the quantitative signals into qualitative signals.

5.4. Brief Review of EMOTION-II Model for Using Emotion as a Feedback to Identify and Correct Unexpected Discrepancy Errors

The EMOTION-II [20] model extends this "emotional feel" model by discovering the discrepancy between the expectancy (internal model prediction by the brain) and the actuality (real world reality) as an emotional feedback for subsequent correction. In this model, emotion is not merely a state of detecting the congruency/incongruency with the reality, but also a process of error recovery in order to reduce the discrepancies. Thus, the emotional resolution process (to become happy) is essentially an error minimization process so that congruency between the expectancy and the actuality is achieved. When this state of congruency is reached, it becomes a state of content or happiness. Conversely, the state of incongruency initiates the emotional feedback that motivates the need for error correction to reduce the incongruency. Thus, emotional resolution, in this model, is essentially an optimization process to reduce incongruency by minimizing the error (the discrepancy between the internal model prediction and the actuality).

The emotional resolution process is essentially a computational process for self-discovery of errors and self-correction of error conditions such that its internal prediction of the external world is congruent with the actual reality. This theory is based on the fact that in order to survive, the internal brain model has to predict the external world accurately. Any error in its prediction could decrease its chance of survival. In order to detect the prediction error autonomously, one of the simplest strategies is to compare the predicted outcomes with the actual outcome in the real world. The difference (disparity error signal) between the expected and the actual outcomes can be used as a measure for subsequent error correction.

Emotion, in this model, is a feedback signal indicating the error condition based on the internally derived disparity signals. Thus, the bigger the disparity, the greater the emotional intensity is, so that it intensifies the emotional response to alert the individual for self-correction. The fault condition would lead to the unhappiness state as a feedback for correction. The congruence between the expected outcomes and the actuality signifies a happy/contented state that does not need correction. Thus, emotion provides a feedback mechanism for self-detection and self-correction of fault conditions for a self-actuating system autonomously. The computational role of emotional response is to use emotion as

a guide in the feedback system for self-adaptive error corrections.

In summary, the EMOTION-II model resolves the fault conditions within the individual itself, *i.e.*, the discrepancy errors within *self*, rather than the discrepancy errors of *other* individuals. The current EMOTION-III model extends the previous model by addressing the fault conditions of the *extended-self*, while including other individuals as a part of the system in the assessment of fault conditions.

5.5. The Empathic Model

This empathic model includes the disparity between self and the other individuals for comparison rather than relying entirely on its internal signals from within oneself for emotional assessment. That is, it incorporates other-regarding concerns into the optimization of self-regarding concerns (as an extended-self) that is necessary in the error-recovery process.

In brief, the previous two emotional models focused on the self-preservation and self-survival as the computational criteria for error-minimization of the desirable outcome (as predicted by the internal model of the external world for an autonomous organism). The present model incorporates the survival of a social group - *i.e.*, the preservation of the whole group (self and others) as an entity - rather than the survival of the individual entity (excluding others).

5.6. Inclusion of Others as Extended-self in Optimization for Survival

In a social group where there is social interaction, the survival problem evolves into an optimization problem of maximizing gains (and minimizing losses) instead of error minimization between the expectancy and actuality. Thus, the probability of survival is increased by maximizing the self-gains, while minimizing the self-losses in social interactions.

Therefore, when each individual attempts to maximize self-gains, it creates the phenomenon of *competition* when the gain for one person could result in a loss for the other person. That is, the process of maximizing self-gains may require maximization of losses for the other person. These social dynamics could result in benefitting one individual, but become detrimental to the other.

This creates the classic "conflict" condition in which the maximization of one variable becomes the minimization of another variable. Therefore, it may not be possible to maximize both variables at the same time, which is the definition of a conflict phenomenon.

5.7. Incorporation of Others as Extended-self Feedback Signals for Optimization

Fig. (**1**) shows the schematic block diagram of the feedback model for optimization for survival. The top half of Fig. (**1**) shows the feedback circuitry for optimization for self-survival, whereas the bottom half shows the feedback circuitry for optimization of the other's survival. Normally, these two individuals are independent of each other in optimizing their own survival. The extended-self feedback signals are not incorporated in the optimization process for self-centered optimization model.
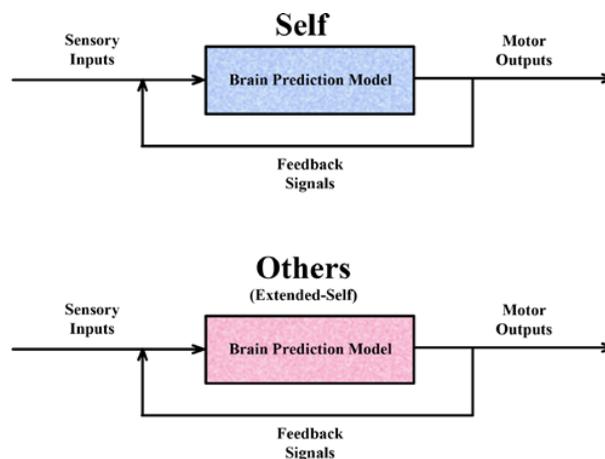
Fig. (1). Schematic block diagram showing the independence of an individual from the other individual without any connection in optimization for survival.

On the other hand, in the extended-self model, the individual is linked to the other individual in the feedback for optimization of survival. By including the other individual as an extended-self, this additional feedback provides the means for empathic emotion formation in maximizing the gain for not just self, but also for others. This linkage is represented schematically in Fig. (**2**) by the block diagram connecting the two individuals together by the feedback signals.

By including the feedback signals from the other individual as a part of its own feedback signals for optimization in survival, the resulting computation would include others in maximizing the gains without excluding the other. Thus, the survival of other is also included in the optimization for survival for one self. This automatically incorporates the other individual as a part of one self. Thus, the ability to feel for others and include others as an integral part of one self creates the empathic emotion for feeling for others and including others as one entity rather than two independent entities. This provides the means for establishing social emotions and cooperative behavior by maximizing the gain for survival for both one self and the other individual without discounting or minimizing the survival of the other.
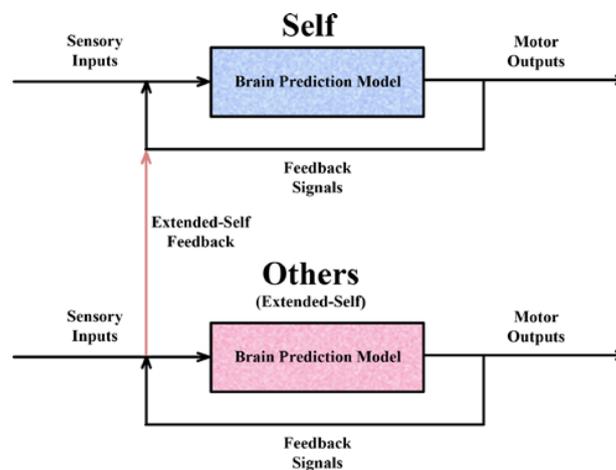
Fig. (2). Schematic block diagram showing the incorporation of others as extended-self for extra feedback in addition to its own feedback signals in forming the empathic emotion. This additional feedback provides the connectivity between two individuals to optimize the survival of both parties rather than discounting or minimizing the survival of the other.

5.8. Optimization Principles in Conflict Resolution

In order to resolve this conflict between two individuals in the optimization process, one of the solutions is to be inclusive of the others in the optimization as if the other person is a part of oneself. It can then resolve the conflict by maximizing the gains for both persons, as if the other person is a part of the self. In other words, it does not require any change in the optimization strategy, except by re-defining the "self" as an "extended self" that includes the other person as a part of self.

By incorporating others in the extended-self model, empathy can be accomplished in the optimization process that minimizes the losses of others rather than maximizing the losses for them. In other words, the survival consideration in the optimization process is not merely considering the survival of self, but the survival of the other also. Thus, this extended-self model provides the theoretical basis of the evolutionary process in social interaction that results in the emergence of empathy as a social emotion. Empathy is emerged when there is an inclusion of others as a part of self in the optimization computation.

When the optimization process extends to include preservation of self and others, it introduces emotional feedback that is not just *self*-regarding, but also *other*-regarding. This provides the basis for social interactions that promote *mutual* survival, and the ability to "feel for others" - *i.e.,* the development of empathic emotions in a social group.

5.9. The Relativistic Model Based on the Frame of Reference

This theoretical model of social emotion is derived based on the ability for an organism to change the frame of reference from *self*-centered to *other*-centered. The change in frame of reference is easily computed by using the relativity principle of directional vectors, which represent the disparity signals between self and others. When the frame of reference is switch from *self* to *others*, the directional vector could change signs - from positive to negative -

resulting in a relativistic change in perception of the comparison between self and other. That is, it changes from a *self*-centered (*self*-regarding) optimization to an *other*-centered (*other*-regarding) optimization. Moreover, the emotional response can be assessed objectively by using a neutral frame of reference that is neither self-centered nor other-centered, but with respect to an independent frame of reference based on a neutral party.

5.10. Mathematical Derivation for Switching Frame of Reference

To illustrate the relativity of a change in the frame of reference, fairness assessment can be used as an example. Fairness is an assessment resulted from comparing the disparity between oneself and others. The disparity can be represented by a signed vector quantity (**d**). Therefore, unfairness is represented by a negative disparity vector (**–d**), while fairness is reflected by the positive disparity vector (**d**) between oneself and others. By switching from a self-centered frame of reference to an other-centered frame of reference, then fairness to oneself becomes unfair to the other person, if the disparity is non-zero.

Then empathy is an emotional response, which can be computed by switching from a self-centered frame of reference to an other-centered frame of reference, such that "perspective switching" can be achieved. The switch in perspective is now viewed from the other person's point-of-view (*i.e.*, "putting in someone's shoes"), resulting in an empathic emotional response as the feedback. Thus, what was considered as unfair earlier would be perceived as fair, when the perspective is changed by switching the frame of reference. Thus, switching the frame of reference is essentially a change the sign of the vector from a negative sign (unfair) to a positive sign (fair) computationally.

5.11. Mathematical Derivation of Empathic Emotion

Let us denote the disparity contributing to the emotional response by a vector, **d**. Let emotion be defined as a *n*-dimensional vector, **e**, which is computed by the vectorial sum (weighted-sum) of all the contributing disparities. It is given by:

$$\mathbf{e} = k_1\mathbf{d}_1 + ... + k_i\mathbf{d}_i + ... + k_n\mathbf{d}_n + b = \sum_{i=1}^{n} k_i\mathbf{d}_i + b \tag{1}$$

where k_i are the weighing factors, and b is a constant (representing the baseline emotion). Without any loss of generality, we can drop the subscript notation, and express the above emotion equation from the perspective of *self*. This is based on the self-centered frame of reference as:

$$\mathbf{e} = \sum_{i=1}^{n} k_i f\left(\mathbf{d}_i\right) + b \tag{2}$$

Let us denote the disparity vector for the other individual using the primed notation, **d′**. Then the emotional vector, **e′**, for the other-centered frame of reference is given by:

$$\mathbf{e}' = \sum_{i=1}^{n} k_i f\left(\mathbf{d}'_i\right) + b' \tag{3}$$

5.12. Change of Perspective in Fairness Assessment

Let us use fairness as an example to demonstrate how a change in the frame of reference can affect the disparity vector to illustrate the relativity of the perspective based on a different frame of reference. Fairness is assessed by a comparison of the disparity between two persons. If two persons have an equal share, then it is usually considered as fair. That is, there is no disparity (**d** = 0) between the two persons. If one person has more than the other person, then there is a disparity between them (**d** ≠ 0). If the disparity is positive (**d** > 0), it is considered as hyper-fair (more than fair) for the person who has more. The disparity vector, **d**, is positive in magnitude based on the self-centered frame of reference. Based on relativity, it would be unfair for the other person (**d′** < 0), and the magnitude of the disparity vector, **d′**, would be negative, based on the other-centered frame of reference.

Therefore, when the perspective is changed from one center frame of reference to another, the sign of the disparity

vector also changes at the same time. This provides the mathematical formulation for the relativity in fairness when a change in the perspective is based on a different frame of reference.

If the emotional response is based on the perception of fairness, then it makes a difference in which center of the frame of reference is used for the assessment of fairness. That is, if a person is happy because this person has more than the other (hyper-fair, $d > 0$), then the other person who has less would be unhappy (because it is unfair, $d' < 0$) [26 - 28, 32]. Thus, the emotional response, in this social context, is relative depending on whose frame of reference is used in the assessment of fairness. It can change from unfair to fair, and consequently from unhappy to happy, by a relativistic change in the frame of reference in assessing the disparity.

5.13. Fairness-equity Graph

In order to illustrate the principle of relativity in fairness perception, the fairness-equity graph can be used to represent the relationship between fairness and equity (or disparity). When the disparity d is a loss ($d < 0$) relative to oneself, then it may seem unfair. Similarly, when the disparity d is a gain ($d > 0$) relative to oneself, then it may seem fair. However, fairness is not necessarily determined by the amount of disparity per se, because the fairness perception can be skewed based on the biases either favoring oneself or the other. Fig. (**3**) illustrates this dichotomy between fairness perception and disparity relativistically.

In Fig. (**3**), the *x*-axis represents the disparity d, and the *y*-axis represents the fairness perception. In the upper-right quadrant, it represents the fairness-equity space that corresponds to hyper-fair and hyper-equitable. That is, because the disparity is a gain for oneself ($d > 0$), it is more than equitable, so the fairness perception would be hyper-fair naturally. Thus, the upper-right quadrant represents the fairness-equity space that is perceived without bias, and it is relative to one's frame of reference rather than the other's frame of reference. It is a gain in this quadrant; therefore, one would feel hyper-fair.
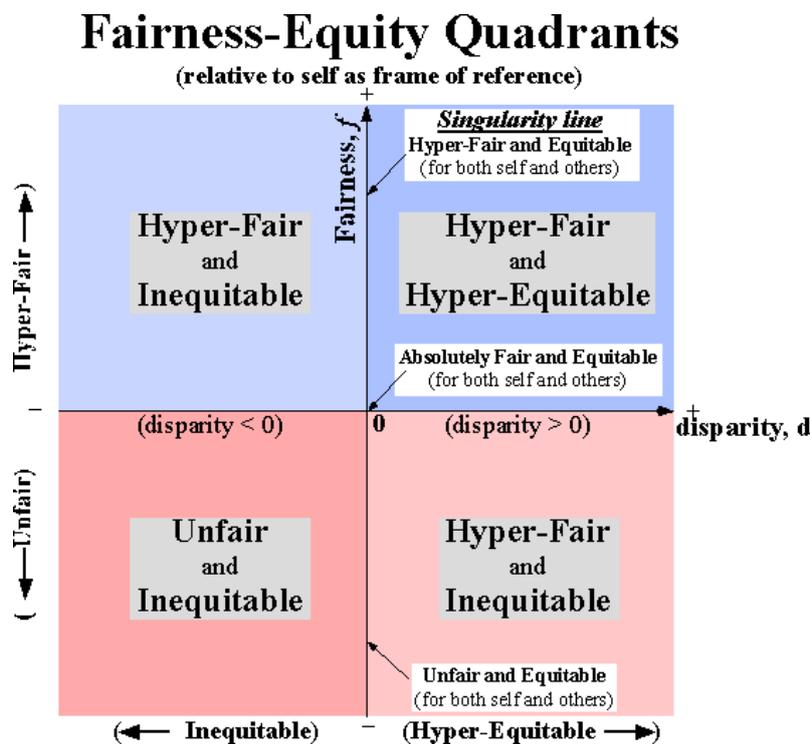
Fig. (3). The relativistic fairness-equity graph representing the relationship between fairness perception and equity (disparity, **d**) relative to oneself. Each quadrant represents the subjective fairness perception with respect to disparity (whether it is inequitable or hyper-equitable).

If the frame of reference were switched from oneself to the other, then the fairness-equity space would move to the lower-left quadrant. In other words, the other's perception of fairness would be unfair and inequitable (*i.e.*, $d < 0$). This

is a loss to the other person; therefore, one would feel unfair. This shows how the relativity of fairness perception can be represented graphically along the diagonal axis, where fairness to one person is unfair to another, and a gain for one person is a loss to another person.

On the other hand, if a person is empathic, and feels for the other person, then the fairness perception can change by moving to the upper-left quadrant. That is, the person would feel hyper-fair even though it is inequitable to oneself ($d <$ 0). The person still feels hyper-fair even though it is a loss rather than a gain because of the ability to feel empathy for the other person, by extending oneself to the other person's perspective. Therefore, this upper-left fairness-equity space represents the empathic quadrant, where a person perceives fairness even though it is inequitable and represents a loss to the self-interest. It is not considered as a loss when the person extends oneself to the other person. When the frame of reference is switched to the other person, the loss for oneself becomes the gain for the other person. Therefore, the fairness perception would switch to hyper-fair instead of unfair. This is the selfless phenomenon in generating the empathy emotion.

Finally, the lower-right quadrant represents the greedy fairness-equity space. It is because even though the disparity is a gain ($d > 0$) for oneself, the person perceives it as unfair. This is the greediness phenomenon that a person is not satisfied with the gain, and perceives the gain as unfair to oneself. Even though this may seem paradoxical in the perception of fairness to feel unfair when a person is gaining more than the other person, it does not violate any logical principle. It is because the fairness-equity threshold is shifted graphically to a lower threshold for fairness. In other words, it requires a large amount of gain in order to be considered as fair. This phenomenon is essentially the greedy feeling while disregarding the other person's perspective. Thus, the fairness-equity graph can be used graphically to represent the relativity of the perception in fairness that results in either empathy or greediness.

Note that the *y*-axis represents the dividing line between the equitable and inequitable quadrants, *i.e.*, no disparity ($d = 0$). This represents the equality between two persons, which is the singularity point (singularity line) in which it is fair for both persons. This is the only condition in which maximizing the gain for both persons can be achieved automatically. Any other space in any of the fairness-equity quadrants is fair for one person but unfair for the other persons relatively.

Thus, egalitarianism is a consequence of the optimization to achieve fairness for both persons, without scarifying either oneself or the other. This resolves the dilemma of maximizing the gain for oneself without maximizing the loss for the other person. This is an example of an empathic response by concerning the other person's fairness in the optimization process, not just considering the fairness based on the self-centered perspective.

5.14. Self-preservation for the Extended Self

When the optimization for survival is taking into account, the self-preservation principle states that the maximization of gains is based on the self-centered perspective. Thus, in a competitive environment where conflict occurs, then fairness for one person becomes unfairness for another person. This occurs when they compete for the same resource, which means that maximizing the gains for one person would consequentially be maximizing the losses for the other person. This is resulted from the relativity principle in the perspective switching. This means that the survival of one person may influence the survival of another person, thus affecting the emotions for both of them, according to this emotional disparity theory in survival.

In order to provide a cooperative interaction where both persons can survive without decreasing the survivability of the other, then the optimization strategy can be changed to extend the *self* into an *extended-self*, which includes the other as a part of oneself. This eliminates the conflict created by optimizing the gains/losses for oneself without regards for the other.

Therefore, by extending the self-centered frame of reference to include the other person computationally, then the new disparity vector, **d** *''*, relative to the extended-self would include both **d** and **d**′, while the sign of the **d**′ vector would not change relatively. That is, the signs of both disparity vectors, **d** and d′, would still be positive, instead of negative for **d**′. It is because the new frame of reference of the other person is now based on the new extended-self centered frame of reference instead of the other-centered frame of reference in the optimization process for the desirable gains. This is the *inclusive principle,* which is the theoretical basis in the formation of empathy by including others as a part of the extended-self in self-preservation and survival.

5.15. Theoretical Derivation of Empathic Emotion Based on the Extended Self Principle

Using the relativity principle of social comparison (*i.e.*, interpersonal comparison), then the disparity vector, $\mathbf{d}$, is positive ($\mathbf{d} > 0$) when it is favorable to self. The disparity vector is negative ($\mathbf{d} < 0$), if it is unfavorable to self in terms of survival. Then the social emotion equation can be given by:

$$\mathbf{e} = \begin{cases} k_{-1}\, f(\mathbf{d}) + k'_{-1}\, f(\mathbf{d}') + b_{-1} & \text{if } \mathbf{d} < 0 \quad \text{(i.e., disparity unfavorable to self)} \\ k_0\, f(\mathbf{d}) + k'_0\, f(\mathbf{d}') + b_0 & \text{if } \mathbf{d} = 0 \quad \text{(i.e., no disparity)} \\ k_1\, f(\mathbf{d}) + k'_1\, f(\mathbf{d}') + b_1 & \text{if } \mathbf{d} > 0 \quad \text{(i.e., disparity favorable to self)} \end{cases} \tag{4}$$

where k_{-1} represents the weighing factor for the unfavorable (inequitable) disparity, k_1 represents the weighing factor for the favorable (hyper-equitable) disparity, k represents the weighing factor for neutral (equitable) disparity, and similar subscripts for the emotional baseline constant.

Therefore, this extended-self model of emotion can account for the empathic emotion - when a person feels for the other person, by including the other person in the computation of the disparity in the optimization process for survival. Thus, this optimization process of the extended self promotes cooperative behavior instead of competitive behavior. It is because if the self-centered frame of reference excludes the other person as a part of the extended-self, then the relativity disparity vector for the other person could change signs. This would result in minimizing the gains for the other person instead of maximizing the gains. Thus, these theoretical derivations provide the computational mechanisms for assessing the empathic emotion for cooperative or competitive behavior in social interaction.

SUMMARY

A theoretical framework of social emotion based on an autonomous control system is derived using a basic set of principles that encapsulate emotions as the emergent properties for increasing the chance of survival in a social group (rather than a single individual). The theoretical framework does not rely on retrospective (or introspective) accounts of psychological experience or artificial construct of what emotions are for, or what the roles of emotions are. Social empathic emotion is derived based on the extended-self model by including others as a part of the extended-self in the optimization process for survival. This extended-self model can incorporate the maximization of self-gains without minimizing the gains for others, even in a competitive environment. This resolves conflicts in social interactions, and promotes cooperative behavior by using empathic emotions as the feedback in the optimization process.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   V. Gallese, "The manifold nature of interpersonal relations: the quest for a common mechanism", *Philos. Trans. R. Soc. Lond. B Biol. Sci.,* vol. 358, no. 1431, pp. 517-528, 2003.
[http://dx.doi.org/10.1098/rstb.2002.1234] [PMID: 12689377]

[2]   S. C. Gadanho, "Learning behavior-selection by emotions and cognition in a multi-goal robot task", *J. Mach. Learn. Res.,* vol. 4, pp. 385-412, 2003.

[3]   S.C. Gadanho, and J. Hallam, *The Role of Emotions Exploring Autonomy Mechanisms in Mobile Robots.,* University of Edinburgh, Department of Artificial Intelligence: Edinburgh, 1997.

[4]   S.C. Gadanho, and J. Hallam, *Emotion-driven Learning for Animat Control.,* University of Edinburgh, Department of Artificial Intelligence: Edinburgh, 1998.

[5]   S.C. Gadanho, and J. Hallam, "Emotion-triggered learning in autonomous robot control", In: *Cybernetics and Systems,* vol. 32. 2001, pp. 531-559.

[6]   A. Takanishi, K. Sato, K. Segawa, H. Takanobu, and H. Miwa, "An anthropomorphic head-eye robot expressing emotions based on equations of emotion", In: *Proceedings of IEEE International Conference on Robotics and Automation,* IEEE: San francisco,CA, 2000, pp. 2243-2249.
[http://dx.doi.org/10.1109/ROBOT.2000.846361]

[7]   M. A. Arbib, "Evolving emotions in animal and robot", *Inter. J. Comp. Intell. Appl.,* vol. 4, pp. 225-236, 2004.

[http://dx.doi.org/10.1142/S1469026804001343]

[8]   R.P. Ebstein, S. Israel, S.H. Chew, S. Zhong, and A. Knafo, "Genetics of human social behavior", *Neuron,* vol. 65, no. 6, pp. 831-844, 2010.
[http://dx.doi.org/10.1016/j.neuron.2010.02.020] [PMID: 20346758]

[9]   F. de Vignemont, and T. Singer, "The empathic brain: how, when and why?", *Trends Cogn. Sci. (Regul. Ed.),* vol. 10, no. 10, pp. 435-441, 2006.
[http://dx.doi.org/10.1016/j.tics.2006.08.008] [PMID: 16949331]

[10]   P.A. Van Lange, "Does empathy trigger only altruistic motivation? How about selflessness or justice?", *Emotion,* vol. 8, no. 6, pp. 766-774, 2008.
[http://dx.doi.org/10.1037/a0013967] [PMID: 19102587]

[11]   J.A. Barraza, and P.J. Zak, "Empathy toward strangers triggers oxytocin release and subsequent generosity", *Ann. N. Y. Acad. Sci.,* vol. 1167, pp. 182-189, 2009.
[http://dx.doi.org/10.1111/j.1749-6632.2009.04504.x] [PMID: 19580564]

[12]   P.J. Zak, A.A. Stanton, and S. Ahmadi, "Oxytocin increases generosity in humans", *PLoS One,* vol. 2, no. 11, p. e1128, 2007.
[http://dx.doi.org/10.1371/journal.pone.0001128] [PMID: 17987115]

[13]   T. Singer, R. Snozzi, G. Bird, P. Petrovic, G. Silani, M. Heinrichs, and R.J. Dolan, "Effects of oxytocin and prosocial behavior on brain responses to direct and vicariously experienced pain", *Emotion,* vol. 8, no. 6, pp. 781-791, 2008.
[http://dx.doi.org/10.1037/a0014195] [PMID: 19102589]

[14]   T. Singer, H.D. Critchley, and K. Preuschoff, "A common role of insula in feelings, empathy and uncertainty", *Trends Cogn. Sci. (Regul. Ed.),* vol. 13, no. 8, pp. 334-340, 2009.
[http://dx.doi.org/10.1016/j.tics.2009.05.001] [PMID: 19643659]

[15]   A. Avenanti, D. Bueti, G. Galati, and S.M. Aglioti, "Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain", *Nat. Neurosci.,* vol. 8, no. 7, pp. 955-960, 2005.
[http://dx.doi.org/10.1038/nn1481] [PMID: 15937484]

[16]   N.D. Tam, "Derivation of the evolution of empathic other-regarding social emotions as compared to non-social self-regarding emotions", *BMC Neurosci.,* vol. 13, p. P28, 2012.
[http://dx.doi.org/10.1186/1471-2202-13-S1-P28]

[17]   D. Tam, "A theoretical model of emotion processing for optimizing the cost function of discrepancy errors between wants and gets", *BMC Neurosci.,* vol. 10, p. P11, 2009.
[http://dx.doi.org/10.1186/1471-2202-10-S1-P11]

[18]   T. Singer, and N. Steinbeis, "Differential roles of fairness and compassion-based motivations for cooperation, defection, and punishment", *Ann. N. Y. Acad. Sci.,* vol. 1167, pp. 41-50, 2009.
[http://dx.doi.org/10.1111/j.1749-6632.2009.04733.x] [PMID: 19580551]

[19]   D. Tam, "EMOTION-I model: A biologically-based theoretical framework for deriving emotional context of sensation in autonomous control systems", *Open Cybern. Sys. J.,* vol. 1, pp. 28-46, 2007.
[http://dx.doi.org/10.2174/1874110X00701010028]

[20]   D. Tam, "EMOTION-II model: A theoretical framework for happy emotion as a self-assessment measure indicating the degree-of-fit (congruency) between the expectancy in subjective and objective realities in autonomous control systems", *Open Cybern. Sys. J.,* vol. 1, pp. 47-60, 2007.
[http://dx.doi.org/10.2174/1874110X00701010047]

[21]   N.D. Tam, "Quantification of happy emotion: Proportionality relationship to gain/loss", *Psychol. Behav. Sci.,* vol. 3, pp. 60-67, 2014.

[22]   N.D. Tam, "Quantification of happy emotion: Dependence on decisions", *Psychol. Behav. Sci.,* vol. 3, pp. 68-74, 2014.

[23]   D.N. Tam, "Computation in emotional processing: quantitative confirmation of proportionality hypothesis for angry unhappy emotional intensity to perceived loss", *Cogn. Comput.,* vol. 3, pp. 394-415, 2011.

[24]   N.D. Tam, "Quantitative assessment of sad emotion", *Psychol. Behav. Sci.,* vol. 4, pp. 36-43, 2015.

[25]   N. D. Tam, and K. M. Smith, "Cognitive computation of jealous emotion", *Psychol. Behav. Sci.,* vol. 3, pp. 1-7, 2014.

[26]   N.D. Tam, "Quantification of fairness perception by including other-regarding concerns using a relativistic fairness-equity model", *Adv. Soc. Sci. Res. J.,* vol. 1, pp. 159-169, 2014.

[27]   N.D. Tam, "Quantification of fairness bias in relation to decisions using a relativistic fairness-equity model", *Adv. Soc. Sci. Res. J.,* vol. 1, pp. 169-178, 2014.

[28]   D.N. Tam, "Quantification of fairness bias by a Fairness-Equity Model", *BMC Neurosci.,* vol. 12, p. P327, 2011.
[http://dx.doi.org/10.1186/1471-2202-12-S1-P327]

[29]   D. Tam, "Variables governing emotion and decision-making: human objectivity underlying its subjective perception", *BMC Neurosci.,* vol. 11, p. P96, 2010.
[http://dx.doi.org/10.1186/1471-2202-11-S1-P96]

[30]   N.D. Tam, "Rational decision-making process choosing fairness over monetary gain as decision criteria", *Psychol. Behav. Sci.,* vol. 3, pp. 16-23, 2014.

[31]   N.D. Tam, "A decision-making phase-space model for fairness assessment", *Psychol. Behav. Sci.,* vol. 3, pp. 8-15, 2014.

[32]   D.N. Tam, "Contributing factors in judgment of fairness by monetary value", *BMC Neurosci.,* vol. 12, p. P329, 2011.
       [http://dx.doi.org/10.1186/1471-2202-12-S1-P329]