(*Preprint prior to published in Cognitive Computation, OnlineFirst™ 12 March 2011*)
http://dx.doi.org/10.1007/s12559-011-9095-2

Computation in Emotional Processing: Quantitative Confirmation of Proportionality Hypothesis for Angry Unhappy Emotional Intensity to Perceived Loss

David Nicoladie Tam

Received: 21 October 2010 / Accepted: 3 February 2011

Abstract A computational model of emotion is derived (using minimalistic assumptions) to quantify how emotions are evolved to estimate the accuracy of an internally generated brain model that predicts the external world. In this model, emotion is an emergent property serving as a self-derived feedback that monitors the accuracy of the internal model via the discrepancy (error measure) between the (internal) subjective reality and (external) objective reality—reality-check subconsciously. Minimization of error (computed by the "gain" toward the desired outcome) will optimize congruency between internal and external worlds—resulting in happy emotion. Unhappy emotion is resulted from the discrepancy between internal and external worlds, which can serve as feedback for self-correction to minimize the "loss" (error) between desired and actual outcomes. Unhappiness provides the internal guide to self-identify whether the cause of error is due to input (sensory perception) error, output (motor execution) error, or modeling (internal model) error. Experimental validation of the hypothesis using the ultimatum game paradigm confirmed the inverse proportional relationship of anger to perceived gain (or direct proportionality to loss) that estimates the discrepancy between what we want and what we get. It also characterizes specific emotional biases by shifting the emotional intensity curve quantitatively.

Keywords Emotional processing • Unhappy • Anger • Fairness • Monetary gain • Ultimatum game • Decision making • Error minimization • Self-discovery of error • Optimization

D. N. Tam
Department of Biological Sciences,
University of North Texas, Denton, TX, USA
email: dtam@unt.edu

1 Introduction

Toward the goal of understanding the computational function of emotion, I have developed a model of emotion that is derived from computational principles in neuro-engineering, such that computational roles of emotion can be assessed objectively with experiments. In order to assess the computational principles of emotion, I first derived a generalizable theoretical framework of emotion, with minimalistic assumptions, using a null hypothesis only (i.e., by not making any a priori assumptions about the purported roles of emotion or the existence of emotion). This reduces any subjective bias that may result from our introspective presumptions of what emotions are used for (the subjective qualia of emotion-related sensations, gut feelings, facial expressions, and their interpretations that may not be generalizable to other animals). In order to reduce any human-centric biases that may interject into the formulation of the computational functions needed for emotional processing, I derive the minimal requirements that are essential for an autonomous system to survive in the real world. Using these minimalistic building blocks, a model is derived using evolutionary principles, such that the universal requirements common for all autonomous beings (animals) can be captured by a theoretical framework. I then show that emotion emerges as one of the basic requirements for increasing survivability, which serves as a feedback for self-discovering any system errors, so that a more accurate estimate of the external world can be achieved. We then verify this hypothetical model with human experiments to confirm whether the hypothesis coincides with reality or reject it if it does not have any relevance in the real world.

(*Preprint prior to published in Cognitive Computation, OnlineFirst™ 12 March 2011*)
http://dx.doi.org/10.1007/s12559-011-9095-2

2 Theoretical Derivations of the EMOTION Model

2.1 Self-bootstrap Method for Increasing Survivability in the Evolution of Emotion by Minimizing Errors

Using a self-bootstrap approach in neurobiological computation, I have identified the minimal requirements for a brain model that increase the chance of survival in animals in the EMOTION-I and EMOTION-II (acronym for "Emotional Model of The Interpretations of Neuroprocessing") models [80, 81]. Essentially, an animal that functions successfully in the environment requires the ability to accurately predict outcomes in the real world by integrating their sensory signals. This task is essentially computed by the internal neural network model in the brain. Because real-world events occur probabilistically in nature, a successful brain model also requires probabilistic predictions. That is, the minimal requirement for an animal to survive successfully is the ability for the internal brain model to produce a probabilistic dynamical model of the external world that accurately predicts the actual reality.

Note that we made no assumptions on the existence of emotion in the derivation at this point, i.e., we only posed the null hypothesis, without hypothesizing any existence or utility of emotions initially. The derivation merely introduces the condition (minimal requirement) for any brain model that survives successfully in the real world is the minimization of its model-prediction errors and other system errors. Whether emotion biases the world-view, or distorts the accurate perception of reality, will be addressed after the derivation of the existence of such emotional biases. It will become apparent later; after we have derived the existence of emotion, the emotional bias is resulted from either distortion or emphasis in the internal brain model, depending on whether such alteration is intentional or not. If the bias is unintentional, it can certainly introduce additional errors to model predictions, compounding the error-correction problem even further. If it is intentionally generated within the system, then it is used to emphasize (or de-emphasize) the neural network (so as to pay attention) to the specific error condition to facilitate the error-recovery process.

Note that error is an intrinsic property of any predictive system. It always exists in any real-world computational models (including robots and animals), independent of whether they (robots or animals) have emotions or whether their emotional biases cause additional errors. In other words, emotional bias of the world-view (distortion of reality) is not a pre-condition or assumption in the derivation of our model, but a side effect of the processing. It is caused by either faulty conditions in the internal neural network model (if it is unintentional) or attentional emphasis (if it is intentional), which we will derive in the following section and also demonstrate the existence of such biasing-effects in the experimental results.

As a preview, we will derive below that emotion is a feedback signal monitoring the animal's error conditions for subsequent corrective actions. Therefore, emotional bias essentially skews the feedback signal in its processing, which creates amplified error conditions in the feedback loop. Furthermore, as a corollary, emotion serves as a feedback not just for "self"-recognition, but also for other animal's recognition (in social communication) to enlist them to assist in this error-corrective process.

2.2 Minimalistic Model Using Minimal Assumptions

The only assumption we made in our model is that the brain has the ability to integrate sensory information to produce motor output by an internal neural network that can learn from its environment. The neural network produces a dynamical time-varying many-to-many mapping function, which maps from the (sensory) input space to the (motor) output space by the following equation:

$$y_j(t) = \prod_{k=1}^{m} \Pr_k \left[\sum_i^n w_{ijk}(t) x_{ik}(t) \right] + a \qquad (1)$$

where $\Pr_k[\bullet]$ denotes a probabilistic nonlinear mapping function, $y_j(t)$ represents the j-th output predicted by the internal model, $x_{ik}(t)$ represents the i-–th input of a neuron at the k-th layer , and $w_{ijk}(t)$ represents the synaptic weight between i-th input and j-th output at the k-th network layer, for a neural network with m layers, and a is a constant. Note that this mapping is a dynamical mapping that encapsulates the interactions between the animal and the environment. The prediction of the neural network is not just a static map of the external world, but a prediction of the dynamical interactions between the animal and its environment (including other animals).

The neural network acquires knowledge of the real world by exploratory action using the Hebbian learning rule [32], which is one of the basic neural mechanisms for learning in animals [7, 14, 15] and in computational neural network models [16, 41, 54–56]. Using such Hebbian associative learning mechanism, time-dependent correlation between the sensory input and

motor output can be established by the correlation function coefficients embedded in the synaptic weights [78, 79]. Applying spike-timing-dependent plasticity rules to the network, the spatiotemporal correlation between sensory input and motor output can be established dynamically [13, 84], such that the model of the external world can be created/ acquired by the internal neural network model through iterative associative reinforcement learning [4, 5, 51, 74, 77].

In the real world, the actual output may not always correspond to the predicted output generated by the internal brain model. The discrepancy between the predicted and actual outcomes represents the error that the brain model needs to correct in order to avoid disastrous results if the animal were to successfully survive. That is, if the internal neural network model accurately predicts the outcomes in the external world, the chance of survival will increase for the animal. If the prediction is inaccurate, its interactions with the external world would become inappropriate and faulty, which often results in a decrease in survivability.

Note that evolution principle tends to produce results that increase the "probability" of survival rather than the "absolute" survivability. Survivability is a measure of resiliency when subjected to the stress-test in competitive and hostile real-world environment. It is the basic principle in evolution – survival-of-the-fittest – where the fittest (most accurate) solution often increases the chance of survival in nature's process of elimination. It never guarantees for survival in the absolute sense because of the unpredictability in nature. Therefore, a global measure of survivability is the increase in long-term survival probability rather than in short-term survival. For instance, sticking the head in the sand is a short-term survival strategy, but it often fails in long-term survival. Thus, ignoring reality (emotional bias or perceptual distortion) may provide a coping strategy for short-term survival, but it often fails in long-term survival or functioning, when reality is distorted (which is typical in most psychiatric/mental disorders, such as anxiety disorders that exaggerate fearful emotion and depression that exaggerates sad emotion). Thus, the emotional bias that serves as a short-term survival coping strategy does not contradict the premise of the present model, because such biasing can facilitate the error-corrective process by emphasizing a specific error condition. But distortion of reality can cumulate enough errors within the system such that it will lead to catastrophic system failure. These pathological conditions often occur in untreated mental illnesses, such as schizophrenia and post-traumatic stress disorder (PTSD).

2.3 Conservative Bias Considerations

Furthermore, an animal that opts for conservative bias to reduce risk in a high-risk environment is a short-term survival strategy that would not contradict the above assertion in long-term survivability. That is because risk corresponds to uncertainty in predicting outcomes with respect to the animal's partial knowledge (i.e., uncertainty in model prediction or reduced accuracy in model predictions). Adopting conservative bias to reduce risk does not necessarily imply improving accuracy in model prediction, because the external risk factors are identical (due to lack of predictive knowledge and/or insufficient sensory data)—independent of whether the animal decides to take any risk or not. The only difference is the animal's decision choice to limit the behavior to a restricted repertoire that is predictable to improve its prediction accuracy within a confined environmental set. That is, it attempts to remain at a local minimum rather than venture out to explore the global minimum. This translates to the well-known optimization problem in the neural network that minimization of errors can arrive at either a local minimum or global minimum [64, 90]. Conservative bias is the skewing of a decision to stop at the local minimum over the decision to explore the possibility of achieving a better global minimum. The risk is not knowing whether a better global minimum exists if one ventures outside its local minimum. Yet, a global minimum often provides a more superior/stable solution than a local minimum [64, 90]. Similarly, long-term survival probability (global minimum) often provides a more stable state than short-term survival (local minimum). Thus, the scenario often corresponds to short-term gain, but long-term loss in the error-minimization process (or a decrease in probability of long-term survival). Such non-risk taking behavior will not pose any contradiction to the premise of the model that an increase in model-prediction accuracy (or reducing system errors) will lead to an increase in the likelihood of long-term survival.

2.4 Accumulation of Internal Errors that Leads to Perceptual Distortion

Note that each of the sensory input and interconnection, $x_{ik}(t)$, at the k-th layer within the neural network can be erroneous, i.e., each contains an error term, $\Delta x^e_{ik}(t)$, such that:

$$x_{ik}(t) = x'_{ik}(t) + \Delta x^e_{ik}(t) \tag{2}$$

where $x'_{ik}(t)$ is the ideal accurate input signal. Similarly, the synaptic weights, $w_{ijk}(t)$, can be

erroneous, containing an error term, $\Delta w^e_{ijk}(t)$, such that:

$$w_{ijk}(t) = w'_{ijk}(t) + \Delta w^e_{ijk}(t) \tag{3}$$

By including the error terms, Eq. 1 (the output of the *j*-th neuron at the *k*-th layer) becomes:

$$y_j(t) = \prod_{k=1}^{m} \Pr_k \left[\sum_{i}^{n} \left[w'_{ijk}(t) + \Delta w^e_{ijk}(t) \right] \left[x'_{ik}(t) + \Delta x^e_{ik}(t) \right] \right] + a \tag{4}$$

which cumulates all the errors from the previous network layers as $\Delta y^e_j(t)$. The sensory errors, $\Delta x^e_{ik}(t)$ can lead to perceptual distortion when such errors accumulate throughout the network by multiplying the sensory errors, $\Delta x^e_{ik}(t)$, with synaptic errors, $\Delta w^e_{ijk}(t)$ in Eq. 4 to produce the cumulative output error, $\Delta y^e_j(t)$.

When errors occur within the processing of the internal neural network, they are extremely difficult to detect. That is because it requires self-examination of the internal processing, but it does not have direct access to the internal interconnects. Most importantly, the neural network would not have knowledge of such internal errors when its knowledge is dependent on the same (potentially faulty) interconnect computation. Yet the discovery of its own internal faults is not necessarily an intractable problem. The neural network can compare the discrepancy between its own prediction with the actual outcome and use it as an indicator to deduce an internal fault condition using this reality-check mechanism.

2.5 Error-Discovery by the Discrepancy Signal between Predicted and Actual Outcomes

Thus, self-discovery of internal fault can be inferred from the discrepancy signal between its own prediction and actual outcome:

$$\Delta y_j(t) = y_j(t) - y'_j(t) \tag{5}$$

where $y_j(t)$ represents the predicted outcome by the neural network (subjective reality), and $y'_j(t)$ represents the actual outcome in the real world (objective reality). A non-zero $\Delta y_j(t)$ represents an inaccurate model, which can be used as an indicator in its self-examination for subsequent error-correction. By minimizing the sum of errors as a measure of the global error-condition, such that

$$\sum_{j}^{l} \left| \Delta y_j(t) \right| \to 0 \tag{6}$$

for all l outputs, then an accurate model of the external world is achieved. If not, the likelihood of survival decreases due to the cumulative errors in predicting the external outcomes accurately.

Note that Eq. 6 varies over time, which means error may be at a minimum at one particular time instance, but will not be a minimum at another time instance, when the animal moves to another environment (which produces entirely different set of environmental conditions for the minimization process). This means that minimization of error is a lifetime process—achieving a minimum global error at one time does not imply that the global error will continue to be a minimum, when the circumstances change at another instance. When we derive (in subsequent sections) that this global error is used as an emotional feedback, this implies that an animal will not stay in an emotional state forever, but the emotional state will change constantly depending on the environmental conditions. This is consistent with the explanation why happiness or unhappiness does not last forever; it changes over time when the circumstances change.

2.6 Error-Discovery by Exploratory Self-Examination

In order to ensure survival, the animal has to assess the accuracy of its own model prediction because there is no presumed a priori knowledge of any unforeseen, unpredictable errors. Because there is no guarantee that its own model is an accurate model, it has to rely on self-derived clues to indicate any possibility of error. This self-derived indictor of error condition is the discrepancy measure of global error given by Eq. 6.

Although potential errors are often unbeknownst to the animal, these errors can be uncovered using the principle of evolution, in which trial and error is used as an ad hoc procedure to sample the solution space by random exploration, i.e., statistical sampling of possible solutions using Monte Carlo simulation or simulated annealing [42, 90, 92]. By iterative exploration of the real world, dynamical prediction of the external world can be approximated by the model prediction as a probabilistic estimate. The accuracy of such a prediction is essential to the survival of the animal, because if the estimation is incorrect, the animal will be more likely to encounter disastrous outcomes when it fails to respond to environmental conditions appropriately.

2.7 Minimal Requirements for Ensuring Survivability: Self-Discovery and Self-Correction of Internal Faults

Given that the minimal survival requirement (for an internal network model) is the ability to self-detect and self-identify errors, then the accuracy of its model prediction can be assessed by the global-error signal. In other words, if an animal is able to self-identify and self-correct errors, then the internal brain model can refine a better estimate of the external world in its prediction. This would result in increasing the likelihood of survival. Therefore, the crucial link to survival is the ability to self-identify and self-correct any error that may exist.

2.8 Autonomous Self-Identification of Error-Sources

Potential errors can come from three major sources:

(a) sensory (input) error, $\Delta x_{ik}^{e}(t)$;
(b) execution (output) error, $\Delta y_{j}^{e}(t)$; and
(c) internal (modeling) error, $\Delta w_{ijk}^{e}(t)$.

But the origin of these errors is often unknown and unpredictable, the animal has to rely on other clues to identify the existence of such errors in order to assess its accuracy. One of the self-derived autonomous schemes for identifying the existence of error is the comparison between model prediction and real world (Eq. 5). Such comparisons between actual reality and the subjective reality (estimated by the brain) can be done autonomously by the neural circuitry within the animal (without needing external assistance from other agents). The discrepancy between the objective reality and internally predicted estimates can serve as the signal for error detection. This provides an autonomous mechanism for self-identification of error for assessing the accuracy of the internal brain model.

2.9 Emotion Hypothesized as Internal Feedback for Self-Discovery of Error-Conditions

This self-derived error signal serves as a feedback to the animal that something is wrong, wherever the source of error may be. We hypothesize that this "error-indicator" coincidentally corresponds to what is commonly known as "emotion" [59]. That is, happiness is a feedback that signifies to the animal that everything is going right as predicted. Unhappiness serves as a feedback that indicates something went wrong, which motivates the animal to correct for such errors. This also corresponds to our intuition to ask what is wrong, when we see someone who is unhappy.

2.10 Proportionality Hypothesis of Emotional-intensity to Discrepancy Error Signal

We also hypothesize that the bigger the discrepancy (error signal), the greater the emotional-intensity will be. This intensifies the motivation to search for solutions to reduce the discrepancy between what it wants the world to be and what happens in actuality. When corrective solution results in the minimization of error, unhappy emotion is resolved. This results in a happy state, and happiness is experienced. Otherwise, the unhappy state remains as a reminding condition for the system to seek corrective solutions either to change the world or to change itself.

Note that corrective action here means reducing system errors by addressing the source of processing errors (faults) within the neural network. The errors to be corrected are conceptual errors or perceptual errors, but not necessarily behavioral errors. Corrective action does not always need alteration of behavioral action, but rather correcting the erroneous processing in these three potential candidate error sources:

(a) input error: faulty perception (for sensory correction), $\Delta x_{ik}^{e}(t)$
(b) output error: faulty execution (for decision correction), $\Delta y_{j}^{e}(t)$ and/or
(c) modeling error: faulty model-prediction (for belief-system correction), $\Delta w_{ijk}^{e}(t)$.

Although in introspection (based on human perspective), anger often motivates reactive action, and sadness immobilizes any behavioral action, they are not the corrective actions that are referred to in our model. The corrective actions are the changes in internal processing—the ways and means to resolve the unhappy emotions by discovering faults (finding out what is wrong), and correcting the error (fixing the perception/decision/belief system) to reach a happiness state (error-free condition). That is, corrective actions are conceptual changes of how the animal perceives the world differently once the false belief is discovered.

2.11 Resolving Emotion by Resolving Faulty Assumptions

Because emotion is an internal feedback in our model, it is often the correction in internal "assumptions" of how an animal perceives the world that resolves the emotion, i.e., changing the world-view. For instance, when a stimulus is no longer perceived as a threat, fearfulness, or anger would be resolved almost instantly. Yet the external circumstances (stimulus conditions) are identical before and after the emotional resolution. The only difference is the internal processing that changed the

faulty perception from a threatening stimulus to non-threatening, or from a faulty belief system believing the stimulus is dangerous to harmless, or from a faulty decision that determines/assesses the stimulus is aversive to benign. Unhappy emotion can be resolved without necessarily any behavioral action, because the corrections are the internal error corrections of the neural network processing in conceptual terms (a change in mind-set) rather than physical terms (even though behavioral changes often occur as a result of such change in belief system).

2.12 Expectancy for the Desirables and Undesirables in Emotions

Expectancy is the prediction generated by the brain. Such prediction may or may not have any behavioral outcome, so it should not be equated to behavioral action. For example, prediction/expectation of reward is an expectancy of pleasurable experience, and prediction of punishment is an expectancy that leads to fear and anxiety. Although these predictions may subsequently lead to seeking behavior (for reward) or avoidance behavior (from punishment), the model predictions are not behavioral actions. The prediction is the subjective reality (believing this is what will happen), and the actual outcome is the objective reality (what happens in actuality, independent of whether the animal believes it is happening or not and denies that it happened or not). Evidence for such internal prediction (in reward) is found in dopamine neurons in the ventral striatum/nucleus accumbens, which are known to encode not only the prediction of reward, but also prediction error [8, 62].

2.13 Expectancy Errors in Predicting the Desirables and Undesirables in Emotions

There are two kinds of expectancy—predicting the desirable (such as reward and pleasure) and predicting the undesirable (such as punishment and pain). This results in two types of expectancy errors. It can evoke a disappointment emotion if the expectation falls short of what it wants or an excitement if it exceeds what it wants. On the other hand, it evokes a frightening emotion if the expected outcome exceeds what it does not want (such as danger, threat, harm, or death) or a reduction in anxiety if the expected outcome is less than the worst-case scenario.

Evidence for such encoding of expectancy errors is numerous. Dopamine neurons in the mesolimbic system is known to be involved in reward-prediction-expectancy, error detection, and prediction error and salience of such reward [8, 10, 37] with distributed processing in orbitofrontal, prefrontal cortex, and basal ganglia for encoding emotional reappraisal of expected value and prediction error [33, 68, 73]. Furthermore, prediction error is encoded as a linear function of reward probability [1], similar to the proportionality hypothesis proposed in our emotional model.

2.14 Wants and Gets in Emotions

When the prediction is what the animal wants (for goal-achievement), it is the "want." When the prediction is what the animal does not want, it is the "unwanted." These are the subjective reality. The "get" is the outcome of what the animal gets, independent of what it wants or does not want. This is the objective reality. The discrepancy error is the difference between the "wants" and "gets"—between the subjective and objective realities—between the prediction and outcome in the reality-check process.

2.15 Difference between Wants-and-Gets and Expectations: the Desirables vs. Undesirables

As described earlier, although expectation is the prediction by the model, such outcome prediction may or may not be desirable with respect to the survival of the animal. For instance, prediction of death (in danger assessment) is an undesirable outcome, unless the animal is suicidal. So accurate prediction of death (such as going over a cliff) would not increase survivability, unless such accurate prediction is used to correct the undesirable outcome (to avoid death by parachuting or bungee jumping). Thus, different emotional feedbacks are evolved to provide the internal tools to differentiate desirable versus undesirable outcome-assessments. For instance, prediction of danger or threat would evoke fear or anger emotion as a feedback to increase survival. Nonetheless, accurate prediction of such danger is essential for survival, because if such prediction is inaccurate, erroneous assessment would lead to inappropriate responses that could mislead the animal into a trap. That is, survivability depends on the accuracy of the prediction of getting the desirables (wants), and avoidance of getting the undesirable (unwanted).

2.16 Survivability Exception: Suicide as a Solution to Resolve the Unresolved Errors

Note that the above derivation is assuming that the system is not suicidal. If so, the desired goal is death; survivability will decrease instead. Nonetheless, this exception will not nullify the hypothesis of our model that unhappy emotion is an indicator for error conditions such that happy state is reached when error reduces. That is because an implicit assumption is made in the derivation that a solution is available to reduce system errors. In the real world, there may not

be any solution, or it gets stuck in finding a viable solution; so error reduction may not be possible. If no apparent solution can be found to reduce errors, then termination of the system could be an alternate solution to completely shut down the nagging unhappy error-feedback signal, so it will never be reminded of the error conditions that it cannot solve. In this case, suicide is a solution to resolve the unresolved emotions, which is often the desired goal of suicidal patients who seek cessation of consciousness as a solution to terminate the unresolved [38, 40, 75]. This is consistent with our hypothesis that emotion serves as an error-feedback signal for the system to correct (assuming such correction is possible) and the process for resolving unhappy emotion is error reduction (error minimization in the optimization process). But if it is not possible to correct such error, termination of the entire system is a solution to terminate the error-feedback signal, thus resolving the unresolved emotions in a radically unconventional manner.

2.17 Happiness as a Congruency Measure between Wants and Gets

Happy emotion serves as a congruency measure for the animal to identify that what it wants is congruent with the real world. When congruency is matched between the wants and gets, it represents a happy state in its internal feedback that the internal model is an accurate model without faults. The model does not need any correction, so it experiences happiness that everything is alright. Reality is exactly what the model predicts. Conversely, inaccurate predictions by the animal will be more likely to decrease its survivability due to faulty estimations.

2.18 Desirability as an Additional Criterion to the Congruency Measure for Emotional Assessment

If an animal gets what it wants (desirable), then happy state is experienced as a feedback for satisfaction. On the other hand, if it gets what it does not want (undesirable), then it would be unhappy instead. For instance, getting an undesirable disease would not be happy, even if such prediction is highly accurate in reality. Therefore, emotion assesses more than the "congruency," but also the "desirability" with respect to survival according to this model. Thus, this survival criterion still fits with the premise of our model that survivability is the key minimal requirement for autonomous beings to function independently. Using self-discover error signal to self-correct internal processing faults completes the feedback loop for autonomous recovery of errors.

2.19 Self-Discovery of Error-Condition as Another Criterion for Emotion

This autonomous self-discovery of an internal fault condition is what distinguishes robots from animals. In fact, this ability for self-discovery of internal error can also be used as one of the criteria to distinguish animals that have emotions versus animals that do not through the subconscious reality-check process.

Evidence for this cognitive emotion regulation by "reappraisal" (reality-check) is encoded in the ventral striatum [33, 72]. This reality-check is also analogous to the discrepancy measure for emotional feedback assessment between the predicted (subjective) reality and the actual outcome (objective reality) proposed in our model. Evidence for these goal-directed behaviors is found to rely on a network that involves the anterior cingulate and prefrontal cortex for error avoidance and reward, which suggests this network's involvement in self-initiated behavioral adjustment, but not necessarily error detection or prediction. In contrast, insula and ventral striatum are more responsive to the high reward expectation [43] and prediction error [1, 62], while the amygdala is involved in re-evaluating the salience value [47] in conjunction with the orbitofrontal cortex to update such re-evaluation [48, 53] for a final decision in choice preference [49].

2.20 Unhappiness as a Discrepancy Measure between Wants and Gets

Conversely, unhappy emotion serves as a discrepancy measure to identify that what it wants is not congruent with the real world. Correction is needed to restore congruency between subjective and objective realities. Furthermore, source of error needs to be identified, if such self-corrective action is effective. Various sub-emotions of unhappiness (such as anger, fear and sadness) coincidentally specifically identify the source of error and corrective action in response to the discrepancy signal. Different emotional reactions (angry, sad and fear) correspond to different coping strategies in the process for self-identification and self-correction of errors.

Once error conditions between the incongruent realities are recognized as not what one wants, then there are two ways to change it—change the world or change yourself. For instance, angry emotion often motivates the animal to change the world (when it does not accept the reality), instead of changing itself (by changing self-perception, world-view, belief system, or decision). Most often, anger is a failed attempt to change the world, because if such attempt were successful, it would no longer be angry when it finally gets what it wants.

2.21 Computational Role of Emotional States for Feedback Processing

When an animal optimizes the system such that error minimization is achieved, it increases the chance of survival as a system. This happy state serves as feedback to the internal model that congruency with the real world is achieved. Otherwise, an unhappy state serves as a guide to indicate the existence of error that needs to be corrected, if happiness is to be achieved. Anger serves as an attempt to change the reality rather than correcting its internal errors in its feedback. Sadness serves as the recognition of loss. Fear serves as the realization that an undesirable prediction threatens its survival. These unhappy emotions identify the specific stages in the error-recognition and error-recovery process via the gain (and loss) computation.

2.22 Quantification of Emotion by Gain and Loss Measures between the Desired and Actual Outcomes

In order to self-assess potential errors, the gain (and loss) measures can provide a quantitative assessment of the difference between expectancy of desirable (predicted reward estimates) and reality (actual outcomes). The gain (and loss) is an internal measure that computes the difference between what an animal (or human) wants and what it gets. If the computed difference between the wants and gets increases, it is a loss measure. If the computed difference decreases, it is a gain measure. Thus, emotional processing depends on the computation of these gain (and loss) measures in order to serve as a feedback to the animal to assess the error conditions.

2.23 Proportionality Hypothesis of Emotion with Gain/Loss Signals

We propose the hypothesis that the emotional intensity is proportional to the gain (and loss) signals computed between the wants (subjective reality) and gets (objective reality). That is, the bigger the gain (or loss), the greater the emotional intensity. The greater the emotional intensity, the greater the feedback conveying the state (or status) of the error conditions to the animal is.

2.24 Neuroengineering Principles in Emotion Processing

Using this neuroengineering scheme, we are able to derive a framework for emotional processing in which quantifiable internally generated variables (error feedback) can be used for emotional computation in order to self-correct any error that may exist in the brain model. This provides the derivation of an emotion model, using basic neurobiological principles, with minimal psychological assumptions about emotion and emotional processing, such that it can be tested by experimental hypotheses.

2.25 Optimization Task in Emotion Processing: Maximization of Gain and Minimization of Loss

The computational task for emotional processing essentially corresponds to an optimization problem that minimizes losses and maximizes gains. It requires self-detection of error derived from the gain (and loss) signals. This results in two distinct classes of processing needed in emotional computation:

(1) maximization of gain for the desirables in happy emotional processing and
(2) minimization of loss (or reducing the undesirables) in unhappy emotional processing.

2.26 Independence of the Maximization and Minimization Processes

Note that the above are two independent processes for optimization even though they serve opposite goals in the computation. That is, the maximization process for happiness does not necessarily imply a minimization for unhappiness or vice versa. That is, one does not replace the other. This is analogous to the opposite effects of the accelerator and brake pedals used to move the car forward or stop it, but taking the foot off the accelerator does not make the car stop nor taking the foot off the brake would not make the car go. Their functions are independent of each other, even though their actions are opposite to each other. This is consistent with the fact that antidepressants are not happy pill, because lifting depression does not necessarily make a person happy. It merely brings a person from an unhappy state back to normal baseline, i.e., reducing the unwanted errors (the undesirables)—minimizing the losses. To be happy, it requires bringing a person above this normal baseline to congruency with the desirables, i.e., maximizing the desirable gains.

2.27 Emotional Bias for Amplifying or Attenuating Importance of Certain Error-Source

The gain (and loss) measures are not necessarily objective measures to the internal brain model:

(1) They can be biased by perceptual error (input error).
(2) They can be skewed by execution or decision error (output error).
(3) They can also be distorted by modeling error (internal prediction error).

Emotional biases can be introduced by scaling the error signal according to the preference of emphasis, i.e., either amplify or attenuate the error signal, as a "error-salience" factor to emphasize/de-emphasize its importance. Thus, a scaling factor (weighing-coefficient) can be applied to Eq. 5, such that Eq. 5 is changed to include the error-salience-bias:

$$\Delta y_j(t) = k\left[y_j(t) - y'_j(t)\right] \tag{7}$$

where k denotes the error-salience weighing-factor or error-salience-bias ($k > 1$ for amplification; $k < 1$ for attenuation). This error-salience-bias may correspond to the valence of emotion or valence effect of prediction in the self-serving bias.

The skewing of these gain/loss estimates essentially amplifies (or attenuates) the error-feedback signal such that emphasis (or de-emphasis) can be placed on these error-sources for further processing. That is, these errors are not treated equally, but are weighted according to the error-salience weighing-coefficient (error-salience-bias) in the weighted-sum summation process. This allows the computation to emphasize or de-emphasize certain error-source in the solution-exploration process. This computational error-salience weighing-factor is an "intentional bias" in the model. This bias factor is different from the "unintentional bias" introduced into the system by faulty signals in faulty perception, $\Delta x^e_{ik}(t)$, faulty execution, $\Delta y^e_j(t)$ and faulty model-prediction, $\Delta w^e_{ijk}(t)$. Emotional bias is the sum of both intentional and unintentional bias, which compounds the problem of whether such bias is essential (or beneficial) to emotional processing or detrimental to the system.

If intentional, emotional bias serves a computational role in selecting which error source is most important to address at the moment. That is because if simultaneous solutions to resolve all errors were not possible, then selecting a specific error may bring attention to the system to address the potential solution one at a time. Therefore, sub-emotions can be used to differentiate/emphasize different solution strategies in resolving error conditions in the emotional computation process. If unintentional, it can lead to pathological conditions.

2.28 Difference between True Emotions in Animals and Error-Correction in Robots

The difference between robots and animals is that true emotions are self-derived internally by autonomous assessment of error conditions without any pre-designed/ pre-programmed error-correction schemes. Most robots often lack this ability to self-derive unknown/unforeseen errors without pre-programming. Moreover, they often do not have the self-conscious awareness process to assess such error conditions.

2.29 Animals with True Emotions vs. Reflexive Actions

As a corollary, we also hypothesize that animals who possess true emotions are those that can self-discover its error conditions for autonomous correction to resolve such errors. Animals who possess pre-programmed error-correction scheme (such as withdrawal reflex from pain to increase survivability) without the ability to self-discover the unforeseen error conditions to change its own internal framework do not necessarily have true emotions, according to this model. That is, it requires self-awareness in the reality-check assessment in emotional recognition to fully autonomously respond to the self-discovered error conditions. Such self-recognition of error conditions is the crucial part in emotional processing, even though this recognition process is subconscious. That is, the identifiable criteria are the ability for the animal to realize when something goes wrong to alert its attention, as well as the ability to recognize the conditions when things go well subconsciously.

2.30 Time-Derivative as an Additional Emotional Measure in Error-Minimization Process to Indicate the Direction Whether Error Increased or Decreased

Since the survival process is an error-minimization process, the system also needs to assess how well the optimization is proceeding. If the discrepancy error increases, the error minimization is heading the wrong direction. If it decreases, it signifies the error correction is proceeding well. Thus, the indicator for assessing the status of the error minimization process is the time derivative of the global error measure:

$$\frac{d}{dt}\left[\sum_j^J \left|\Delta y_j(t)\right|\right] = \frac{d}{dt}\left[\sum_j^J \left|y_j(t) - y'_j(t)\right|\right] \tag{8}$$

which indicates whether the global error increases or decreases depending on the sign of this time-derivative.

This time derivative of emotion provides the transient response—surprise signal—such as excitement in happy emotion and shock in unhappy emotion, when the unexpected happens. This emotional derivative serves as a feedback signal to indicate whether the optimization process is homing in the correct direction. If the gain is not one expected (such as gaining a disease), then the emotional

feedback would be unhappy, which indicates that error becomes larger than expected, i.e., the optimization is heading in the wrong direction.

Because Eq. 8 is a time-dependent (time-varying) measure for transient response, it becomes zero if nothing changes (i.e., if the error remains constant). This accounts for the phenomenon that transient excitement emotions often fade away as time passes. That is, ecstasy does not last forever, but satisfaction, or contented state of happiness, can last because the happy state is a state of the system corresponding to the global error, rather than a transient measure, which corresponds to the rate of change of error.

To revive the transient response, new wants and gets are often needed. Thus, the system is always seeking improvements to get more and more of what it wants. This motivates the animal to seek and explore better and better solutions (to reach global minimum), if the current solution is at a local minimum.

2.31 Second-Order Time-Derivative as Another Additional Emotional Measure in Error-Minimization Process to Indicate Minimum, Maximum or Point-of-Inflection

In order to assess whether it arrives at a minimum or maximum, second-order time derivative of Eq. 8 is needed:

$$\frac{d^2}{dt^2}\left[\sum_j^I \left|\Delta y_j(t)\right|\right] = \frac{d^2}{dt^2}\left[\sum_j^I \left|y_j(t) - y'_j(t)\right|\right] \qquad (9)$$

This provides additional information on whether such error reached a minimum or a maximum, or point of inflection (a transient point), not just merely whether the error correction is heading the right direction. This allows the system to recognize whether it reached a stable point or an unstable point, in the optimization process.

3 Experimental Confirmation of the EMOTION Model

In order to verify the hypothesis in our EMOTION, we employed the ultimatum game (UG) paradigm to elicit potential emotions in human subjects so that we can assess the emotional response with respect to the desired gains and losses. UG is a classical paradigm widely used in behavioral economics for assessing decision making in neuroscience, psychology, social science, economics, and mathematical psychology [9, 36, 70, 89]. It is an experimental paradigm in which an amount of money (such as $10) is divided between two persons (a proposer and a responder). If the responder accepts the offer, both keep the money. If the responder rejects it, both lose the money. Rather than addressing decision-making process as in most traditional UG analyses [17, 61, 66, 67, 71, 91], we use this paradigm to provide a means for subjects to self-generate endogenous emotions. We can then evaluate the gain/loss disparity with respect to whatever emotional response the subjects produced.

3.1 Experimental Null Hypothesis

In order to verify our emotional response hypothesis, it is essential to test the null hypothesis in which no emotion responses are assumed in the experimental design; otherwise, it would fall trap into proving the self-fulfilling prophecy if we assume the subjects will respond to any specific emotion. Therefore, we specifically choose this UG paradigm (split-the-money game) because the stimulus condition is merely a neutral stimulus of monetary offer, without any emotional content. We do not manipulate any of the psychological conditions of the subjects or present any stimulus that would evoke any intentional or unintentional responses, other than the monetary offer to accept or reject. The subjects are free to feel any emotion (or no emotion) to the monetary offer, without any manipulation by the experimenter to skew their perception or responses.

3.2 No Manipulation of Experimental Conditions

By design, we do not want to control or influence any assumptions the subjects may make about who the proposer is. The experiment is done online via computer terminal without any hint of human versus computer (without any indication of whether the proposer is a human or computer), because it is known that subjects do respond differently to computer proposer compared with human proposer [61, 67] if such hints were made. No human interactions or suggestions (such as pictorial clues, facial expression, or gender of the proposer) were used. The survey questionnaire was presented completely in written text form (without any pictorial clues). The responders answer the questions by a mouse-click to the answers.

3.3 Neutrality and Objectivity of Experimental Conditions

It is imperative for us to present the experimental conditions as neutral as possible across subjects to prevent any procedural biases (without any hints or suggestions to the subjects that may influence their mind-set, decision or emotional responses). Any emotion experienced by the subjects would be entirely self-induced (rather than suggested/aroused by the experimenter), depending on the subject's own perception of fairness or any other assumptions that the

subject may have in splitting the money. They are free to decide to accept or reject the offer by making any assumption about the offer they wish, without any coercion/ suggestion by the experimenter or experimental design.

3.4 No Assumption of Any Specific Emotional Response

Although it is widely known that decision to accept or reject the offer often depends on the emotional response to fairness, empathy, altruism, or social norms [6, 17, 27, 31, 44, 45, 60, 63, 66, 67], we choose not to make any of these assumptions in our experimental design to avoid experimenter bias.

4 Methods

4.1 Ultimatum Game Paradigm Splitting $10

We collected data from 425 subjects (age ranging from 18 to 80, median = 21; 275 women, 150 men) from a pool of metadata from multiple experiments sharing the same UG design of splitting $10, which is used as the baseline control experiment for comparison to the various subsequent experiments, since splitting $10 is the most commonly used paradigm in most published UG studies. Randomized one-shot trial offers (without repeating) between $1 and $9 were proposed to the subjects. That is, the experiment trials consist of proposing 9 pseudo-random offers—ranging from $1 to $9—for the subjects to accept or reject (by clicking the accept or reject button in the computer screen with the mouse). By design, we use the same pseudo-random sequence uniformly for all subjects, so that we can compare the longitudinal practice effects and desensitization effects across subjects (Tam, in preparation). We also survey their emotional state before and after the experiment to verify the test–retest reliability, and to document how emotion changes over the course of experiment.

4.2 Self-Reported Emotional Rating Scale for Cognitive Self-Assessment of Emotion

Self-reported emotional ratings (?5 to -5 scale) are recorded for each trial after they accept (or reject) the offer (by clicking one of the buttons in the rating scale with the mouse). No time restrictions were imposed to complete the survey, nor did we record the timing information of their responses because we do not want to introduce any unintentional restrictions that may affect/perturb the subject's response into making impulsive decision that could affect the emotional state because of frustration or time-pressure. Timing information is essential in fMRI studies (in our subsequent studies) for synchronizing between stimulus and emotional response in order to correlate the emotion with neural responses, but since we are not recording any neural or physiological responses in the current design, the timing parameters are non-essential in the present study.

We use self-reported rating of emotion to assess the cognitive assessment of their own emotion because our objective is to measure their emotion cognition rather than the internal neural/physiological emotional states. This allows us to deduce the relationship between the cognitive response and hidden variables involved in emotional processing to verify our emotional-processing hypothesis before correlating them with neural responses using fMRI recordings later. Although the self-reported cognitive response is often filtered/biased by subjectivity, our goal is precisely measuring how such subjective biases are reported by the subjects, which will reveal the conditions under which emotions are skewed/altered using our analysis. Because the self-reported emotions are merely "cognitive assessment of the subject's own emotion," for brevity, we abbreviate this as "emotional response."

4.3 Multiple Emotions and Distractors as Controls

In order not to skew the subject's response into biasing their self-report favoring any particular emotion, we present the subject with a list of emotions (happy, sad, angry, and jealousy) to rate, in addition to a list of other distractors (by asking how important winning, money, and fairness is for them to rate, using the same ?5 to -5 scale; how fair the offer was; and whether they won that trial or not). This ensures the neutrality and objectivity of the questionnaire in the experimental design to minimize any unintentional side effects that may contaminate the subject's perception or assumption of what they think the experimenter wants from them.

The analysis in this paper focuses primarily on the angry emotion to be concise. The study was conducted with the university Institutional Review Board approval. Informed consents are provided to subjects prior to the experiments.

5 Results

5.1 Emotional-Intensity as an Inverse Proportional Relationship with Gain-Ratio

Figure 1 displays the self-reported rating of angry emotion independent of whether they accept or reject the offer for the entire sampled population. The data approximate an inverse proportionality relationship with monetary gain ratios and anger (Fig. 1a), which is non-random ($r^2 = 0.872$; $p<0.001$). They reported not-

angry for all offers—although much less angry for generous offers than stingy offers.

Note that the singleton point ($5:$5) deviates from the overall proportionality trend, indicating that subjects respond to the even-split ($5:$5) much different than the rest of the fair or hyper-fair offers. This singleton deviation is much more acute in our fairness analysis [88], in which subjects identify even-split ($5:$5) as the fairest based on an objective frame of reference, while the other offer-ratios are perceived subjectively relative to their own frame of reference [83]. That is, they do identify and differentiate the difference between objective and subjective fairness. This objectivity also correlates with their reported objective emotional response to $5:$5 offer, which is different from subjective emotional response to the other offers.

5.2 Emotional-Intensity Conditioned on Fairness (Generous Offers) vs. Unfairness (Stingy Offers)

To reveal whether such proportionality exists depending on their perception of fairness, we subdivide the same data into two subsets in the curve-fitting algorithm (Fig. 1b)— the unfair trials (\$5:$5, left-half) versus the hyper-fair trials ([$5:$5, right-half). A distinctly different proportionality relationship is revealed, with a discrete change of emotional sensitivity from stingy (unfair) to generous (hyper-fair) offers and a singleton point in the middle fair offer (even-split of $5:$5, middle). That is, the emotional proportionality relationship is different for unfair versus hyper-fair trials, even though overall anger relationship is still proportional to the stinginess of the offer (monetary offer-ratio). It shows emotional-intensity changes depending on whether they see the offer as fair or not.

Note that in terms of monetary gain, generous offers ([$5:$5) are often considered as hyper-fair, while stingy offers (\$5:$5) are considered as unfair [66, 67]. However, fairness ratio is not always centered on $5:$5, but skewed toward stingy offer or generous offer, dependent on the subject's subjective bias in their perception of fairness [28– 30, 83]. In order to simplify our analysis, ignoring the subjective fairness bias for a moment (assuming fairness is centered on $5:$5) and assuming money and fairness are the desired outcomes wanted by the subjects, and then unfair offers create the classical dilemma—getting only money or fairness, but not both. This allows us to differentiate the discrepancy between wants and gets in the gain/ loss measures with respect to money and/or fairness. We will focus our analysis on emotions generated by subjects in response to the amount of gain/loss, rather than analyze whether their decisions are rational or not [17, 39, 52].

5.3 Emotional-Intensity Conditioned on Acceptance Decision

To identify the factors that may affect the emotional response, we separate the subject's response conditioned on their acceptance (Fig. 2) versus rejection decisions (Fig. 3). The emotional baseline changes drastically, i.e., they get angrier when they reject the offers. Although the overall inverse proportionality relationship remains in both cases, the emotional-intensity curve of anger shifts upward from not-angry in acceptance trials (Fig. 2a) to angry in rejection trials (Fig. 3a). They reported not-angry for all monetary offers when they accept the offers (Fig. 2), independent whether it is fair or not. That is, when they get (the money) they want, they are not-angry; which is consistent with our hypothesis of getting what one wants in happy emotion.

5.4 Emotional-Intensity Conditioned on Rejection Decision

When they rejected the offers (Fig. 3), they consciously acknowledged their anger (left-half of Fig. 3a, b) when the offers are unfair, but self-reported not-angry when the offers are hyper-fair (right-half of Fig. 3a, b). Interestingly, subjects do not just reject unfair offers. They do reject hyper-fair offers also, for whatever reasons, but not out of anger, because they reported not-angry when they reject such generous offers. It is not a denial, because they do report angry when rejecting unfair offers. If they were in denial of anger, they would not have reported their emotional intensity of anger as a continuous function (straight line) proportional to the stinginess of the offer (or inversely proportional to the gain ratio) (Fig. 3a). That is, the rejection is only emotionally driven by their anger when the offer was stingy (unfair) and but not when the offer is generous (fair or hyper-fair).

Their decision to reject offers is not necessarily emotionally driven by anger (as most UG studies assumed), as subjects reported not-angry when they turn down generous offers (probably because money is not what they want, they are not starving for a few dollars, or if they regard the offer as bribery or sweetening deal for them). In fact, they reject more on hyper-fair offers than $5:$5 even-split perfectly-fair offer, which suggests that they may be suspicious of the ulterior motive of the proposer to offer such generous offer, and reject it due to mistrust, don't need or want any money, if they were rich, or other legitimate reasons rather than reject it out of anger or

irrationally. This is consistent with our hypothesis that they would not be unhappy (or angry) if money is not what they want, when rejecting the generous offers (Fig. 3a, b, right-half). In other words, if they see it as dirty money, they would be happy to reject it, even if the offer is extremely generous [86]. They are unhappy only when they do not get what they want (losing both money and fairness when they rejected unfair offers), which is exactly the emotion they reported in rejecting the stingy offer (Fig. 3a, b, left-half). This is consistent with the negativity bias that people tend to attribute proportional losses (misfortunes) to intentional agent subjectively, but objectively neutral to gains (fortunes) [46]. Last but not least, they probably reject the offer because of their bias in fairness perception, when they actually self-reported perceiving hyper-fair offers as unfair, shifting the fairness-curve down to unfair [88].

Thus, the decision to reject monetary offer can be very rational and is consistent with our emotional hypothesis for reducing the discrepancy between wants and gets. If no such discrepancy exists, they would not experience unhappiness, which is what they reported in rejecting the generous offer.

5.5 Non-Randomness of Emotional Responses

This also shows that subjects took the experiment seriously, and did not reject the monetary offers randomly, or out of no good reasons. It also shows that subjects did not deny the emotions in their self-report, or left their emotion unreported, when they accurately reported angry when rejected unfair offers, but not-angry when they rejected hyper-fair offers or when they accepted offers of any monetary amount. When the offers were proposed in random order, they would not have consciously reported the proportional emotional intensity in a sequential manner according to the descrambled monetary gain ratio as plotted in the graphs. Most importantly, these graphs represent the average response from the entire sampled population such that even if individual subjects respond differently and feel differently emotionally (which they do), as a whole, they do respond proportionally to the gain ratio in a consistent manner. Regardless of individual variability, the graphs reveal the proportional relationship in the analysis.

5.6 Persistence of Proportionality Relationship Conditioned on Decision

Independent of the rationale behind responder's decision, the data demonstrated the proportionality of emotional intensity associates with the perceived gain/loss, regardless of their decision to accept or reject the offers (Fig. 1a). If their decisions were taken into account, the proportionality relationship still remains (Figs. 2a, 3a). This demonstrates the robustness of the proportionality relationship in our emotional-intensity hypothesis.

If decisions were taken into account, the baseline of the emotional-intensity curve up shifts upward from not-angry to angry—in the decision process from accepting to rejecting the offer, regardless of whether the offer is fair or not, stingy or generous (Figs. 2a, 3a). That is, the intercept moves up. This corresponds to the baseline emotion changes from not-angry to angry by 2.5 points in a 5-point scale (i.e., 50% of the angry-scale or not-angry-scale). Whether emotion alters the decision, or decision alters the emotion, is yet to be determined. Nonetheless, a correlation between decision and emotion is quantified by the emotional-intensity curves here.

Interestingly, this drastic shift in baseline emotion according to their decision is independent of their fairness perception. That is, the emotional baseline changes by half-of-its-max-intensity if the decision is changed from accepting to rejecting, for both hyper-fair and unfair trials, but there is only an emotional sensitivity change from hyper-fair and unfair. In other words, the emotional baseline is dependent on decision more than fairness. This is consistent with the finding that anger, spite or revenge is a bigger factor than fairness perception in their decision to reject [58].

5.7 Emotional-Intensity Conditioned on Fairness (Generous Offers) vs. Unfairness (Stingy Offers)

In order to identify the condition in which fairness may affect their emotion, we separate the curve-fitting into hyper-fair and unfair groups (Figs. 1b, 2b, 3b). When fairness (offer-ratio) is taken into account, it reveals a change in the slope of the emotional-intensity curves—a change in emotional sensitivity. That is, each increment of $1 is perceived differently emotionally, depending on whether it is fair or not. This sensitivity change is different for acceptance versus rejection decision (Figs. 2b, 3b). In other words, the emotional intensity to fairness is interrelated to decision, rather than independent of decision. More specifically, when they decide to accept the money, the emotional intensity is rather constant for unfair offers (Fig. 2b, left-half). This suggests as long as they got the money, they are rather indifferent to the amount of unfairness, i.e., the exact amount of unfairness does not change their emotion by much. The emotion does change proportionally when it is a hyper-fair offer (Fig. 2b, right-half). On the other hand, when they decide to reject the money, the opposite is true. Their emotional intensity is indifferent to hyper-fairness; that is, they do

not care how generous the offer is, it will not change their heart when they decided to reject it. But if the offer is unfair, and if they decide to reject it, then their anger increases with the stinginess of the money-offer ratio.

5.8 Shifting of Emotional-Intensity Curve Upward from Acceptance to Rejection Decision

Taken the above two dependence factors together, the emotional-intensity curve is shifted/altered more by decision than fairness consideration. In other words, decision is a bigger hidden factor that biases angry emotion to a different level than the fairness hidden factor, or that anger is a bigger factor than fairness in biasing their decision. On the other hand, fairness affects the emotional sensitivity on the perception of disparity in the gain ratio. Although the present experimental design does not allow us to differentiate whether decision actually biases emotion or emotion biases decision, (or whether fairness biases emotion or emotion biases fairness perception), our results did reveal the hidden factor that affects (correlates with) emotional bias is greater for decision than fairness.

The finding that decision has a much greater effect on emotion than fairness is very much consistent with our discrepancy hypothesis of emotion measuring the outcomes to assess congruency with reality. It is the decision that creates the outcome of reality whether the responder actually gets the money or not. Without such decision, no reality-check be possible. Only after a decision is made, will the reality-check be possible to compare the difference between the predicted and actuality. This provides the emotional feedback to assess whether such decision is a wise decision to resolve the discrepancy between what they want and get and whether it actually reduced the discrepancy. In case of rejecting the money, it failed to resolve the discrepancy (when they realize they lost the money); thus, it amplifies the emotion further for getting nothing.

5.9 Proportional Continuum of Emotional-Intensity from Acceptance to Rejection

One of the most unexpected results from the analysis is that when Fig. 2a is combined with Fig. 3a side-by-side (by appending Fig. 2a to the right of Fig. 3a), a continuous straight line can be connected from rejection trials to acceptance trials (Fig. 4). This suggests that the decision to accept or reject is highly correlated with the emotional intensity in a continuum. As anger intensity decreases from angry to not-angry, the decision changes from rejection to acceptance. The cutoff point for this switch in decision (from rejection to acceptance) is not centered on the neutral point of anger (zero angry-intensity rating at $5:$5 offer-ratio in Fig. 3a), but at the -1.5 angry-intensity (at $10:$0 for rejection and $0:$10 for acceptance, center of Fig. 4). That is, for those who rejected the offer, if the proposer had offered them all $10, they will change their mind and accept it. For those who accepted the offer, if the proposer had offered them no money ($0), they would reject it instead of accept it. This suggests they may have already made up their mind before the offer was given.

5.10 Emotional Bias in Shifting Baseline Emotional Threshold for Switching Decision

This shows the decision to accept or reject is a continuous function of emotional intensity. The switching condition to change from rejection to acceptance is not pivoted at neutral (zero) anger threshold level, but shifted by -1.5 point (30% of 5-point scale) in the emotional-intensity scale. That is, small amount of happiness (not-angry) would not change their mind. It takes a minimum threshold of 30% happiness (30% not-angry) to accept the offer; or that it takes 100% of the monetary offer-ratio to convince those who reject to accept the offer, or 0% of the offer-ratio to convince those who accept to reject the offer. This shows the emotional bias in the threshold shift in emotional baseline for switching their decision.

Because we cannot differentiate whether emotional response leads to decision or decision leads to emotional response using this paradigm, what we can deduce is that decision and emotions are highly correlated in a continuum, in spite of the various hidden factors (such as fairness) that slightly alter such continuity into step-functions (Fig. 5) rather than a straight line (Fig. 4). That is, the underlying predominant factor correlated with decision is emotional level, while fairness merely alters the sensitivity factor at a particular emotional level. This is consistent with the earlier report that anger is a bigger contributing factor than fairness in rejection decision [58] and extends the earlier finding beyond anger into non-anger in rejecting hyper-fair offer because of the downward shift in baseline angry-emotional threshold for decision to switch. Most importantly, Fig. 5 shows anger emotional intensity is not only a function of offer-ratio (monetary gain-ratio), but also depending on whether the offer-ratio is greater than one ([1) or less than one (\1). The brain somehow computes the gain-ratio differently with different bias (skewing effect) to produce different reported emotional level.

5.11 Cognitive Emotional Response After the Decision

Note that the emotion they reported are recorded after the decision had made, in which the reported emotion is more likely to be a resolved emotion as a consequence of their decision than a transient emotion in response to the offer prior to their decision. Due to the limitations of the experimental design, we cannot differentiate which emotion they are reporting to.

Nonetheless, this shows their desire for fairness and their desire for monetary gain can influence on their emotional response from angry (if they want both, but end up getting none, when they rejected the money)—to not-angry (if they want one or the other, and end up getting some, when they accept the money). This is consistent with our hypothesis that emotion resolution is a process for optimizing the system by error reduction (reducing the discrepancy between wants and gets, by checking with reality). Such discrepancy can be reduced by minimizing the wants (wanting less) or maximizing the gets (getting more), which is consistent with the report that satisficers (who opt for fewer choices) are happier than maximizers (who opt for more choices) [20].

The discrepancy is biggest—when they want (both money and fairness), but get none (no money nor fairness)—which is reflected in their reported anger (Fig. 3, left-half). The discrepancy is the least—when they want less (choose money or fairness, but not both), but getting more (at least getting some money)—which is reflected in their reported happiness [86] and not-angry (Fig. 2).

Whether such decision is rational or not is not an issue here, as far as the emotional feedback is concerned. The feedback exactly indicates they got nothing in reality, a big discrepancy from what they want in their expectation. Even though we did not measure (or would not know in any way, given the experimental design) exactly what they want in their mind, it can be deduced from the analysis that this is one of the likely scenarios. Similarly, whether the motive for their anger is revenge or altruistic punishment in rejecting falls into the theory-of-mind ([76], [91]) in social reciprocity that our experimental design did not address, nor is it part of our current emotion model or hypothesis to verify.

Nonetheless, if revenge were a motive for rejection, it would be consistent with our anger hypothesis that it is a failed attempt to change the world rather than correcting its own error. Such failed attempt would appear as irrational because it did not accomplish the intended goal to regain fairness by rejecting the free money. The finding that altruistic punishment is more associated with anger than unfairness [69] is consistent with our model that anger motivates behavior directed externally to others (rather than internally) in the attempt to resolve the unexpected difference between what one wants and gets.

We will address such emotional feedback for social communication in our next EMOTION-III social interaction model (Tam, in preparation). This next model will be the natural progression from the evolution of emotional "feel" in sensation (i.e., pleasant/unpleasant sensation) in the EMOTION-I model as the building block to the evolution of emotional error-discovery feedback in EMOTION-II model. It will extend these two prior "self"-recognition models to "others"-recognition model to provide error feedback for social interaction in real-world environment with other autonomous agents (animals and humans) to involve them in the error-correction process.

5.12 Footnote on Regression Coefficient

As a footnote, the regression coefficient r^2 value is dependent on (1) the slope, (2) residual errors, and (3) the range of x, the correlating variable. That is, r^2 increases as the slope increases – which represents the hypothesis that the variables x and y are related. If the slope is zero, $r^2 = 0$ also, which means the putative variables x and y are not related, and they are independent of each other. But r^2 also decreases as the residual error increases in the curve-fitting. Therefore, caution has to be taken to interpret small r^2 values – it could mean a shallow slope or the range of x is small, but not necessarily large residual error (poor goodness-of-fit). To account for the difference in r^2 values in Panel A vs. Panel B, let us consider the above 3 factors that affect r^2 one by one:

(1) the reduction in r^2 values is due to smaller range of x used:
   (a) in Panel A: a range of 9 (from \$1:\$9 to \$9:\$1)
   (b) in Panel B: a range of only 4 (from \$1:\$9 to \$4:\$6) for left-half, a range of only 4 (from \$6:\$4 to \$9:\$1) for right-half;

(2) the reduction in r^2 values is due to shallower slope (independent relationship between x and y);

(3) the reduction in r^2 values is due to an increase in residual error, which is caused by the reduction in statistical power and variability in smaller sample size (n). This is evidenced by the large standard-error-of-mean (SEM) in

right-half of Fig. 3 (which represents only a small number of subjects rejected hyper-fair offer), when compared to the small SEM in right-half of Fig. 2 (which represents large number of people who accepted hyper-fair offers).

6 Discussion

The results are consistent with our hypothesis that unhappy (angry, in this case) emotional intensity is directly proportional to loss of desirables (or inversely proportional to the desired gains). The proportionality relationships can change, depending on the perceived fairness and acceptance/rejection decisions. That is, based on the graphs, the emotional curves are shifted according to the subjective biases by the subjects. The intercept of the emotional-intensity curve (baseline emotion) changes depending on decision, whereas the slope of the emotional-intensity curve (emotional sensitivity) changes depending on the perceived fairness. Note that there is a discrete shift from the hyper-fair trials to the unfair trials, indicating the sudden change in anger response when encountered with fair versus unfair offers, but that shift is not as great as the difference between acceptance and rejection decisions.

6.1 Empirical Emotional-Intensity Curve Established

The subjective emotional biases can be accounted for by two objective parameters (slope, w, and intercept, b) in the emotional curves by the empirical emotional-intensity equation, similar to the happy emotional curve [86] and jealousy emotional curve [87], except for the difference in proportionality relationship:

$$E = wG + b \tag{10}$$

where E denotes emotional-intensity, G denotes the (monetary) gain ($G > 0$, represents gain, and $G < 0$ represents loss), w represents the weighing-factor for scaling the emotional sensitivity (slope in the graph), and b corresponds to the residual baseline emotion (intercept in the graph). The weighing-factor w alters the emotional sensitivity to a particular gain/loss, i.e., the difference in emotional response for each $1-increment loss. It corresponds to the amplification of emotion bias to increase its sensitivity for selecting which of the error-measures would be optimized first, as described in our model earlier. The residual emotion b represents the baseline emotion.

Note that Eq. 10 is an empirical equation derived from the experimental data rather than derived from our theoretical model earlier. This is why we did not introduce this equation in the theoretical derivation section because this equation is derived from the data analysis rather than from the model. This shows the convergence of evidence on the quantification of emotional bias based on independent derivation from two separate approaches—different computational and experimental methodologies. Yet they both arrive at the same conclusion demonstrating the congruency of our emotional bias hypothesis theoretically and experimentally.

This residual term, b, in Eq. 10 can further be generalized to include the sum of all other gains (or losses), not just monetary gain/loss G, but also other factors, such as fairness, decision, gender [85], and any other unaccounted hidden factors. Therefore, the residual emotions, b, can be represented by:

$$b = w_1X_1 + w_2X_2 + w_3X_3 + ... + c \tag{11}$$

where w_i represents the weighing-factor for scaling emotional sensitivity for each gain factors X_1, X_2, X_3, etc. and c represents the constant corresponding to the intrinsic (innate) baseline emotion. Generalizing G as one of the generalized variables X_1 (i.e., $G = X_1$), Eq. 10 becomes:

$$E = \sum_{i=1}^{n} w_i X_i + c \tag{12}$$

This generalized emotional-intensity equation, E, can be represented by a multi-dimensional function where X_1, X_2, …, X_n are the independent variables in a n-dimensional graph. In this UG experimental example, X_1, represents the monetary gain, X_2 represents the gain in fairness, and X_3 represents the decision.

6.2 Quantification of Emotional Biases by Shifting of the Emotional-Intensity Curve

The cognitive self-reported emotional intensity is skewed (biased) by shifting the emotional-intensity curve (up or down, left or right), or changing the slope or intercept, or the shape of the proportionality relationship. In other words, emotional bias can be quantitatively by:

(a) lifting/depressing the emotion – shifting curve up/down,
(b) biasing the perception (perceived gains/losses) – shifting curve left/right,
(c) altering the emotional sensitivity – changing the slope, and/or

(d) exaggerating the emotion—changing from a linear proportionality (Figs. 2a, 3a, 4) to other non-linear functions (Figs. 2b, 3b, 5).

There are many other hidden factors, which are known to contribute to the rejection and emotional response in other UG studies. Ventromedial prefrontal cortex (vmPFC) damage can cause an increase in rejection rate to unfair offers in UG [39]. Testosterone level in men can increase the rejection rate in UG [12], yet, in the contrary, testosterone in women can actually increase prosocial cooperative fair bargaining behavior and reduce conflicts compared to placebo in UG [23]. Low level of serum serotonin can also increase rejection rate in UG [18, 19, 24] Low level of serum omega-3 fatty acids, which is linked to impulsivity and hostility, can also increase rejection rate in UG [25]. Even sadness induced by watching a sad movie can bias the UG decision [31]. However, it is not known how these hidden factors affect the emotional-intensity level in UG, or they merely affect/impair the decision judgment unrelated to emotions.

6.3 Skewing in Emotional Processing

Thus, emotional processing involves not only computation of the gain/loss signals, but also the skewing effect of such computation via shifting of the emotional-intensity curve. The shifting of emotional-intensity curve captures the subjective biases of emotion graphically by the mathematical function in Eq. 12. The equation provides an intuitive assessment of the richness of emotional response simply by shifting the emotional-intensity curve representing the qualitative distortion (subjective bias) of emotion by humans.

Given that emotional bias can enhance emotional processing by selecting specific error signal for error reduction, small emotional bias can be advantageous to survival. But if such emotional bias becomes too exaggerated, it could lead to distortion of reality, resulting in pathological conditions, such as affective disorders and other psychiatric dysfunctions, caused by distortion of emotional perception or exaggerated emotional responses. With a theoretical framework for emotion established, systematic quantification of emotion can be done experimentally to identify the various hidden factors that affect such emotional processing.

6.4 Rationality in Rejecting Monetary Offers in Ultimatum Game

It may not be a paradox or irrationality when a person rejects monetary offers in UG. Just because it is "free" money, they do not have to accept it, if money is not what they want. The rational monetary maximization assumption (as reported in most UG studies) presumes if they were rational, they would always accept any offer because it is free money, and irrational if they reject it. This assumption does not always hold true, not just in UG settings, but also in other real-life scenarios where no rational human (or animal) would always accept sex offer just because it is free. They have to want it to accept it. In fact, it is very repulsive to accept unwanted sex, which would lead to an unhappy state, as predicted in our model, when a person gets something that he/she does not want. It is very rational to reject unwanted offers.

This paradox is analogous to the cultural legend that "you can bring a horse to the water, but you cannot make him drink." The horse needs to be thirsty before he will want to drink. To assume the horse will accept any water, just because we offer them free water, is just an irrationality of human assumption that we imposed onto the horse in this cultural myth, rather than the irrationality or the stubbornness of the horse when he refuses it. This is because of the biological evidence that animal behavior is often driven by physiological needs, and by extension, psycho-physiological wants. It is a simplistic assumption to assume any rational human will accept any free money handed out to them, without consideration of other factors. Decision always requires selection of choices, conditioned on the circumstances. If the choice is fixed regardless of the circumstances, it is a reflex rather than a decision.

The rationality for decisions to reject (or accept) is also consistent with Nash equilibrium in behavioral economics [9, 50] where the optimizing variable includes not just money, but other variables, such as reward and punishment [70], fairness [52], and social reciprocity, such as altruism [65] and vengeance [58], intentional choice of proposer [26], and other cost-functions of value [82]. The rational decision to reject monetary offer is similar to the decision in behavioral economics to pay a higher price for a product (rather than the cheapest price), when the optimizing variable is not just price, but also quality of the product.

Furthermore, the above emotional responses reported by the subjects are consistent with the "Maximization Paradox" in "Paradox of Choice." "Maximizers" who opt for more choices often end up feeling less satisfied (when they get less than what they want) while "satisficers" who opt for less choices often end up feeling more satisfied (when they get exactly what they want) [20].

In addition, the "reinforcer devaluation effect" could also have played a role here; when a person (or

animal) is satisfied with the reward, subsequent identical rewards are devaluated, i.e., they worth less once they are satisfied [2, 3, 21], in which basolateral amygdala is crucial in the expression of reinforcer devaluation [35].

The conjecture that any true rational responder should accept any non-zero amount of money, according to rational maximization of payoff, is often an assumption based on self-interest of monetary gain only, but even a simple mathematical models (with minimalist assumptions without any reasoning or intelligence capacity) predicts fairness can be evolved out of interactions in UG to reject unfair offers rather than accepting any free money merely for monetary gains [22, 57] by reaching a dynamical equilibrium at the fairness point (by computing the differential cost-functions) without needing any intelligence or reasoning ability. In fact, the report that chimpanzees were rational maximizers in UG when they did not to reject offers [34] was due to the delayed reinforcement in their experimental protocol rather than their inability to evaluate fairness. When humans were asked to delay their rejection response from 1 min to 5 min, humans too acted like the reported chimpanzees as rational maximizers of payoff not to reject any offers due to the cost of delayed reinforcement to forgo the rejection and accept the offers [67]. Furthermore, capuchin monkeys were shown to reject unequal pay using an alternate experimental paradigm [11].

This shows humans and other animals do incorporate other cost factors (including reinforcing rate, consideration of others and fairness) in their decision, which can override the monetary gain or the cost of loss in their decision. Furthermore, the rejection rate in UG changes depending on the knowledge of the availability of alternate offers for the proposer to offer to the responder ([26], [29]), which suggests the responder incorporates the intention of the proposer as well as outcomes in their decision.

6.5 Rejection of Null Hypotheses

Although impulsivity may play a role in shaping the decision to reject remunerative, but unfair, offers, our current experimental paradigm is not designed to address the reasons why they make such decision (as in most other UG studies), but rather the emotional correlates with the gain/loss measures to prove our emotion hypothesis. In fact, as stated in our null hypothesis, we do not want to assume any preconceived notions of what emotions are used for or how emotions bias decisions to prevent proving a self-fulfilling prophecy. As a corollary, we apply similar null hypothesis for the decision-making process, i.e., by assuming the decision is random, rather than determined by some preconceived factors, such as anger or impulsivity. That is why our experimental design specifically avoids coercing the subjects to make their decision in a certain way, timing their response to force them to make an impulsive decision, or coaxing them to feel any particular emotion. This allows us to reveal the consistency of our results demonstrating a proportionality relationship, in spite of the potential random variability introduced by the uncontrolled and unaccounted hidden variables if the null hypothesis were true. Thus, we can reject the null hypothesis that emotional intensity is unrelated to gain/loss, and the null hypothesis that their decision is random and unrelated to emotion and fairness is unrelated to emotion, when consistent and non-random results are correlated with the putative variables in our hypothesis.

6.6 Consistency of Gain-Ratios with the Emotional Correlates

Specifically, the proportionality relationship emerges when gain ratio is correlated with the angry emotion variable (Fig. 1a), regardless of the hidden variables of fairness and decision. When fairness hidden variable is taken into account alone (Fig. 1b), the proportionality relationship persists for both hyper-fair and unfair conditions, albeit slightly different proportionality. When decision hidden variable is taken into account alone (Figs. 2a, 3a), the proportionality also remains for both acceptance and rejection conditions, although shifted upward. When both decision and fairness hidden-variables are taken into account simultaneously (Figs. 2b, 3b), the proportionality still exists, but changes into a step-function. That is, no matter how many hidden variables we included in the analysis, the emotional-intensity proportionality relationship with gain ratio exists in all cases, in addition to further revealing how such these hidden variables alter the proportionality relationships. Most strikingly in our analysis is that the emotional-intensity relationship is a continuum that spans across decision choices (Fig. 4), which changes into a step-function if fairness is taken into account (Fig. 5).

Furthermore, similar proportionality relationships were demonstrated quantitatively for other emotions—happy emotion [86], jealousy emotion [87], sad emotion (Tam, in preparation), and fairness [83, 88]). These proportionality relationships also exist when gender is taken into account, when other monetary values ($10-split vs. $10 million-split) are taken into account and when other quality values (money vs. love) are taken into account [85]; in preparation). The

linearity relationship is also consistent with the finding that similar prediction error is encoded as a linear function of reward probability by neural recordings of the mesolimbic neurons [1].

The robustness of the relationship between emotional intensity and the gain/loss ratio is consistent across all hidden variables that we had examined that it is sufficiently convincing to reject the null hypothesis and accept the hypothesis that the putative variable of discrepancy error is indeed an emotion-related variable in the emotional processing loop.

6.7 Experimental Caveats

Note that due to the design of the experimental survey, the emotional responses reported by the subjects were arrived at the time after they made the decision to the proposal, rather than before or during their decision was made. This means the emotion they reported could be an emotional resolution after the decision rather than the transient emotion during their emotional processing prior to the decision. As discussed in the theoretical derivation section, the transient emotions are different from the resulting emotional states. Although no timing constraints were imposed on the subjects, most subjects made the decision and reported their emotions within seconds after the proposal was offer. So the reported emotional response can represent a mix of transient emotions and stabilized emotional state of resolution. Regardless of this variability and uncertainty of what stage the emotional processing they are reporting, the results did show consistent pattern among a large population of subjects.

In order to delineate the exact neural emotional processing, time-sequence of transient emotional response before, during, and after the decision from the time of proposal offer will be recorded using simultaneous fMRI and EEG recordings in subsequent studies. This will allow us to determine whether such proportionality relationship is preserved at the neural coding level and how it varies with the emotional processing stages.

Furthermore, the results reported here are the emotional states reported through the psychological filters of the subjects, including the possibility that they could be in emotional denial. If they were in denial, the results would not be expected as consistent as revealed in our analysis. Nonetheless, brain imaging and electrophysiological recordings of neural firings will provide direct measures of whether such filtering, biasing or denial may have occurred that are undetectable due to the limitations of the present experimental methodology.

7 Conclusions

The experimental data collected from human subjects confirmed the hypothesis of the EMOTION models that unhappy emotion is inversely proportional to the gain between desired and actual outcomes. The data also quantified the skewing (biasing) effect of other hidden factors (such as perceived fairness and decision and other cognitive and subconscious factors) by changing the slope and intercept of the empirical emotional-intensity curve. The data show that decision shifts the emotional-intensity baseline up/down (by changing the intercept), while fairness changes the emotional sensitivity (by changing the slope). It shows that decision is a much bigger factor on biasing emotional level than fairness, or vice versa.

Most interestingly, emotional intensity can be shown to be a continuum from acceptance to rejection decision. It is a straight-line linear function if fairness is not taken into account. If fairness is taken into account, this linear function changes into a step function. That is, anger emotional intensity level is computed not only based on the gain-ratio, but also skewed depending on whether gain-ratio is greater than one (>1) or less than one (<1), confirming the dependence of emotional intensity on gain/loss measure as predicted by our model.

This demonstrates that emotions are quantifiable measures used for computation within the brain to process error conditions in the optimization process to minimize self-discovered errors by a subconscious reality-checking process, as predicted by the theoretical model. By establishing a comprehensive theoretical framework addressing the computation of emotions, it will facilitate the design of future experiments to reveal how emotion is biased by hidden variables associated with emotional processing in both normal cognitive functions and pathological dysfunctions in affective disorders.

Acknowledgments I appreciate the comments and suggestions by the anonymous reviewers. I also thank Richelle Trube and Krista Smith for proofreading the manuscript.

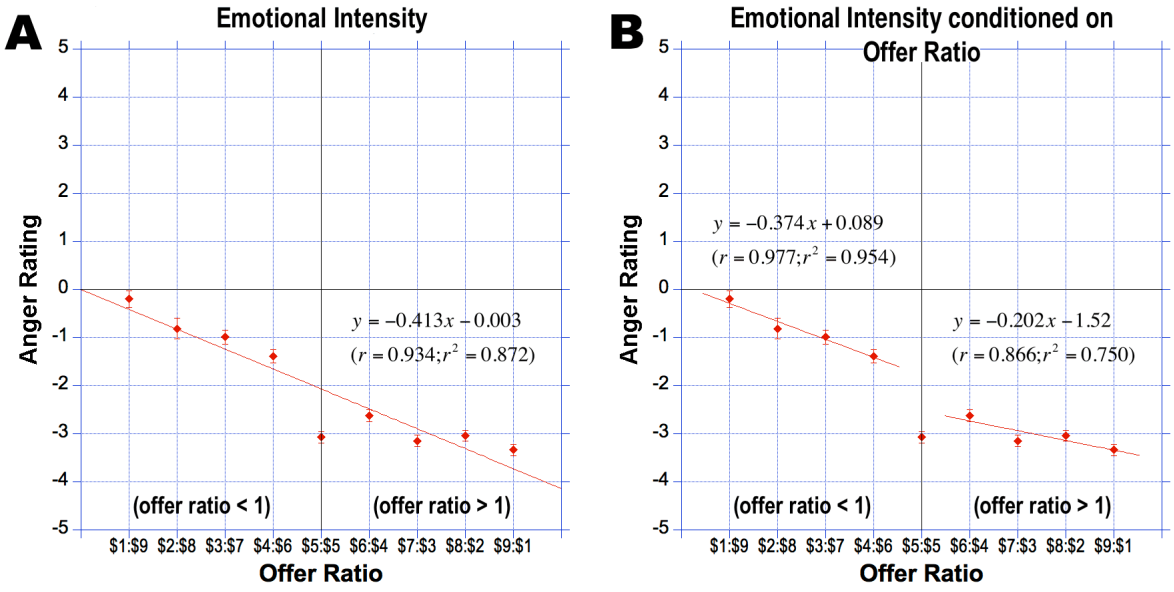
References

1. Abler B, Walter H, Erk S, Kammerer H, Spitzer M. Prediction error as a linear function of reward probability is coded in human nucleus accumbens. Neuroimage. 2006;31:790–5.
2. Adams CD. Variations in the sensitivity of instrumental responding to reinforcer devaluation. Q J Exp Psychol. 1982;34B:77–98.

3. Adams CD, Dickinson A. Instrumental responding following reinforcer devaluation. Q J Exp Psychol. 1981;33B:109–21.
4. Barto AG, Sutton RS. Landmark learning: an illustration of associative search. Biol Cybern. 1981;42:1–8.
5. Barto AG, Anderson CW, Sutton RS. Synthesis of non-linear control surfaces by a layered associative search network. Biol Cybern. 1982;43:175–85.
6. Bechara A. The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. Brain Cogn. 2004;55:30–40.
7. Bender VA, Feldman DE. A dynamic spatial gradient of Hebbian learning in dendrites. Neuron. 2006;51:153–5.
8. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl). 2007;191:391–431.
9. Braun DA, Ortega PA, Wolpert DM. Nash equilibria in multi-agent motor interactions. PLoS Comput Biol. 2009;5:e1000468.
10. Bray S, O'Doherty J. Neural coding of reward-prediction error signals during classical conditioning with attractive faces. J Neurophysiol. 2007;97:3036–45.
11. Brosnan SF, De Waal FB. Monkeys reject unequal pay. Nature. 2003;425:297–9.
12. Burnham TC. High-testosterone men reject low ultimatum game offers. Proc Biol Sci. 2007;274:2327–30.
13. Bush D, Philippides A, Husbands P, O'Shea M. Spike-timing dependent plasticity and the cognitive map. Front Comput Neurosci. 2010;15:142.
14. Butz M, Wörgötter F, van Ooyen A. Activity-dependent structural plasticity. Brain Res Rev. 2009;60:287–305.
15. Caporale N, Dan Y. Spike timing-dependent plasticity: a Hebbian learning rule. Annu Rev Neurosci. 2008;31:25–46.
16. Chauvin Y. Principal component analysis by gradient descent on a constrained linear Hebbian cell. In: Proceedings of IJCNN, Washington, vol. I. 1989. p. 373–80.
17. Civai C, Corradi-Dell'Acqua C, Gamer M, Rumiati RI. Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. Cognition. 2010;114:89–95.
18. Crockett MJ. The neurochemistry of fairness: clarifying the link between serotonin and prosocial behavior. Ann NY Acad Sci. 2009;1167:76–86.
19. Crockett MJ, Clark L, Tabibnia G, Lieberman MD, Robbins TW. Serotonin modulates behavioral reactions to unfairness. Science. 2008;320:1739.
20. Dar-Nimrod I, Rawn CD, Lehman DR, Schwartz B. The maximization paradox: the costs of seeking alternatives. Pers Individ Differ. 2009;46:631–5.
21. Dickinson A, Nicholas DJ, Adams CD. The effect of instrumental training contingency on susceptibility to reinforcer devaluation. Q J Exp Psychol. 1983;35B:35–51.
22. Duan WQ, Stanley HE. Fairness emergence from zero-intelligence agents. Phys Rev E Stat Nonlinear Soft Matter Phys. 2010;81:026104.
23. Eisenegger C, Naef M, Snozzi R, Heinrichs M, Fehr E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. Nature. 2010;463:356–9.
24. Emanuele E, Brondino N, Bertona M, Re S, Geroldi D. Relationship between platelet serotonin content and rejections of unfair offers in the ultimatum game. Neurosci Lett. 2008;437:158–61.
25. Emanuele E, Brondino N, Re S, Bertona M, Geroldi D. Serum omega-3 fatty acids are associated with ultimatum bargaining behavior. Physiol Behav. 2009;96:180–3.
26. Falk A, Fehr E, Fuschbacher U. On the nature of fair behavior. Econ Inquiry. 2003;41:20–6.
27. Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgment. Neuron. 2004;44:389–400.
28. Güroğlu B, van den Bos W, Rombouts SA, Crone EA. Unfair? It depends: neural correlates of fairness in social context. Soc Cogn Affect Neurosci. 2010 (Advance Access published March 28, 2010).
29. Güroğlu B, van den Bos W, Crone EA. Fairness considerations: increasing understanding of intentionality during adolescence. J Exp Child Psychol. 2009;104:398–409.
30. Halko ML, Hlushchuk Y, Hari R, Schu ¨rmann M. Competing with peers: mentalizing-related brain activity reflects what is at stake. Neuroimage. 2009;46:542–8.
31. Harlé KM, Sanfey AG. Incidental sadness biases social economic decisions in the Ultimatum Game. Emotion. 2007;7:876–81.
32. Hebb DO. The organization of behavior. New York: Wiley; 1949.
33. Herwig U, Baumgartner T, Kaffenberger T, Brühl A, Kottlow M, Schreiter-Gasser U, Abler B, JänckeL, Rufer M. Modulation of anticipatory emotion and perception processing by cognitive control. Neuroimage. 2007;37:652–62.
34. Jensen K, Call J, Tomasselo M. Chimpanzees are rational maximizers in an ultimatum game. Nature. 2007;318:107–9.
35. Johnson AW, Gallagher M, Holland PC. The basolateral amygdala is critical to the expression of Pavlovian and instrumental outcome-specific reinforcer devaluation effects. J Neurosci. 2009;29:696–704.
36. Kagel JH, Roth AE. The handbook of experimental economics. Princeton: Princeton Univ Press; 1995.
37. Khamassi M, Mulder AB, Tabuchi E, Douchamps V, Wiener SI. Anticipatory reward signals in ventral striatal neurons of behaving rats. Eur J Neurosci. 2008;28:1849–66.

38. Kienhorst IC, De Wilde EJ, Diekstra RF, Wolters WH. Adolescents' image of their suicide attempt. J Am Acad Child Adolesc Psychiatry. 1995;34:623–8.
39. Koenigs M, Tranel D. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. J Neurosci. 2007;27:951–6.
40. Kraft TL, Jobes DA, Lineberry TW, Conrad A, Kung S. Brief report: why suicide? Perceptions of suicidal inpatients and reflections of clinical researchers. Arch Suicide Res. 2010;14:375–82.
41. Krogh A, Hertz J. Hebbian learning of principal components. In: Eckmiller R, Hartmann G, Hauske G, editors. Parallel processing in neural systems and computers. Amsterdam: Elsevier; 1990. p. 183–6.
42. Ma W, Yu C, Zhang W. Monte Carlo simulation of early molecular evolution in the RNA World. Biosystems. 2007;90:28–39.
43. Magno E, Simo ˜es-Franklin C, Robertson IH, Garavan H. The role of the dorsal anterior cingulate in evaluating behavior for achieving gains and avoiding losses. J Cogn Neurosci. 2009;21:2328–42.
44. McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. Science. 2004;306:503–7.
45. Miller E, Cohen J. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001;24:167–202.
46. Morewedge CK. Negativity bias in attribution of external agency. J Exp Psychol Gen. 2009;138:535–545.
47. Morrison SE, Salzman CD. Re-valuing the amygdala. Curr Opin Neurobiol. 2010;20:221–30.
48. Murray EA, Izquierdo A. Orbitofrontal cortex and amygdala contributions to affect and action in primates. Ann NY Acad Sci. 2007;1121:273–96.
49. Murray EA, Wise SP. Interactions between orbital prefrontal cortex and amygdala: advanced cognition, learned responses and instinctive behaviors. Curr Opin Neurobiol. 2010;20:212–20.
50. Nash J. Essays on game theory. Cheltenham: Elgar; 1996.
51. Niv Y. Reinforcement learning in the brain. J Math Psychol. 2009;53:139–54.
52. Nowak MA, Page KM, Sigmund K. Fairness versus reason in the ultimatum game. Science. 2000;289:1773–5.
53. O'Doherty JP. Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. Ann NY Acad Sci. 2007;1121:254–72.
54. Oja E. A simplified neuron model as a principal components analyzer. J Math Biol. 1982;15:267–73.
55. Oja E. Principal components, minor components, and linear neural networks. Neural Netw. 1992;5:927–36.
56. Oja E, Ogawa H, Wangviwattana J. Learning in non-linear constrained Hebbian networks. In: Kohonen T, Mikisara K, Simula O, Kangas J, editors. Artificial neural networks. Amsterdam: North-Holland; 1991. p. 385–90.
57. Page KM, Nowak MA. A generalized adaptive dynamics framework can describe the evolutionary Ultimatum Game. J Theor Biol. 2001;209:173–9.
58. Pillutla MM, Murnighan JK. Unfairness, anger, and spite: emotional rejections of ultimatum offers. Organ Behav Hum Decis Process. 1996;68:208–24.
59. Plato. The republic (trans: Jowett B). 360B.C.E. http://www.gutenberg.org/ebooks/1497.
60. Quirk GJ, Beer JS. Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. Curr Opin Neurobiol. 2006;16:723–7.
61. Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. The neural correlates of theory of mind within interpersonal interactions. Neuroimage. 2004;22(4):1694–703.
62. Rodriguez PF, Aron AR, Poldrack RA. Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning. Hum Brain Mapp. 2006;27:306–13.
63. Rolls ET. Brain mechanisms of emotion and decision-making. Int Congr Ser. 2006;1291:3–13.
64. Rumelhart DE, McClelland JL, The PDP Research Group. Parallel distributed processing—vol 1, Foundations. Cambridge: MIT Press; 1986.
65. Sánchez A, Cuesta JA. Altruism may arise from individual selection. J Theor Biol. 2005;235:233–40.
66. Sanfey AG, Loewenstein G, McClure SM, Cohen JD. Neuroeconomics: cross-currents in research on decision-making. Trends Cogn Sci. 2006;10:108–16.
67. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of economic decision-making in the Ultimatum Game. Science. 2003;300:1755–8.
68. Schultz W, Tremblay L, Hollerman JR. Reward processing in primate orbitofrontal cortex and basal ganglia. Cereb Cortex. 2000;10:272–84.
69. Seip EC, van Dijk WW, Rotteveel M. On hotheads and Dirty Harries: the primacy of anger in altruistic punishment. Ann N Y Acad Sci. 2009;1167:190–6.
70. Sigmund K, Hauert C, Nowak MA. Reward and punishment. PNAS. 2001;98:10757–62.
71. Smith P, Silberberg A. Rational maximizing by humans (Homo sapiens) in an ultimatum game. Anim Cogn. 2010;13:671–7.
72. Staudinger MR, Erk S, Abler B, Walter H. Cognitive reappraisal modulates expected value and prediction error encoding in the ventral striatum. Neuroimage. 2009;47:713–21.
73. Stefani MR, Moghaddam B. Rule learning and reward contingency are associated with dissociable patterns of dopamine activation in the rat prefrontal cortex, nucleus accumbens, and dorsal striatum. J Neurosci. 2006;26:8810–9918.
74. Sutton RS, Barto AG. Toward a modern theory of adaptive networks: expectation and prediction. Psychol Rev. 1981;88:135–70.

75. Szanto K, Gildengers A, Mulsant BH, Brown G, Alexopoulos GS, Reynolds CF III. Identification of suicidal ideation and prevention of suicidal behaviour in the elderly. Drugs Aging. 2002;19:11–24.
76. Takagishi H, Kameshima S, Schug J, Koizumi M, Yamagishi T. Theory of mind enhances preference for fairness. J Exp Child Psychol. 2009;105:130–7.
77. Tam DC. A positive/negative reinforcement learning model for associative search network. In: Shirazi B, editor. Proceedings of the 1st annual IEEE symposium on parallel and distributed processing. 1989. p. 300–7.
78. Tam DC. Computation of cross-correlation function by a time-delayed neural network. In: Dagli CH, Burke LI, Ferna ´ndez BR, Ghosh J, editors. Intelligent engineering systems through artificial neural networks, vol. 3. New York: American Society of Mechanical Engineers Press; 1993. p. 51–5.
79. Tam D. Theoretical analysis of cross-correlation of time-series signals computed by a time-delayed Hebbian associative learning neural network. Open Cybern Syst J. 2007;1:1–4.
80. Tam D. EMOTION-I model: a biologically-based theoretical framework for deriving emotional context of sensation in autonomous control systems. Open Cybern Syst J. 2007;1:28–46.
81. Tam D. 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 Syst J. 2007;1:47–60.
82. Tam D. A theoretical model of emotion processing for optimizing the cost function of discrepancy errors between wants and gets. BMC Neuroscience. 2009;10(Suppl 1):P11.
83. Tam D. Variables governing emotion and decision-making: human objectivity underlying its subjective perception. BMC Neurosci. 2010;11(Suppl 1):P96.
84. Tam D. Temporal associative memory (TAM) by spike-timing dependent plasticity. BMC Neurosci. 2010;11(Suppl 1):P105.
85. Tam D. Gender difference in emotional perception of love in a decision-making task. Program No. 307.19. Neuroscience Meeting Planner. San Diego: Society for Neuroscience; 2010c (online).
86. Tam D. Cognitive perception of happy emotion: proportionality relationships with gains and losses when getting what one wants; 2011 (submitted).
87. Tam D. Cognitive computation of jealousy emotion: inverse proportionality relationships with gains/losses when one wants something that one cannot get; 2011 (submitted).
88. Tam D. Objectivity in subjective perception of fairness: relativity in proportionality relationship with equity by switching frame of reference –- a fairness-equity model; 2011 (submitted).
89. Von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton: Princeton University Press; 1953.
90. Wu S, Chow TW. Self-organizing and self-evolving neurons: a new neural network for optimization. IEEE Trans Neural Netw. 2007;18:385–96.
91. Yamagishi T, Horita Y, Takagishi H, Shinada M, Tanida S, Cook KS. The private rejection of unfair offers and emotional commitment. Proc Natl Acad Sci. 2009;106:11520–3.
92. Zhang SQ, Ching WK, Ng MK, Akutsu T. Simulation study in Probabilistic Boolean Network models for genetic regulatory networks. Int J Data Min Bioinform. 2007;1:217–40.



(*Preprint prior to published in Cognitive Computation, OnlineFirst™ 12 March 2011*)
http://dx.doi.org/10.1007/s12559-011-9095-2

Fig. 1 Average response of entire population (n = 425) for self-reported emotional rating of angry emotion with respect to monetary offer-ratio (regardless of acceptance or rejection decision). **a** Curve-fitting to the entire sample shows a linear inverse proportional relationship between anger and monetary gain (or direct proportional to loss). **b** It shows same data as in a except the curve-fitting is conditioned on hyper-fair (generous) versus unfair (stingy) trials. Different inverse proportionality relationships appear depending on the perceived fairness (or unfairness) of the offers (Error bar represents SEM)

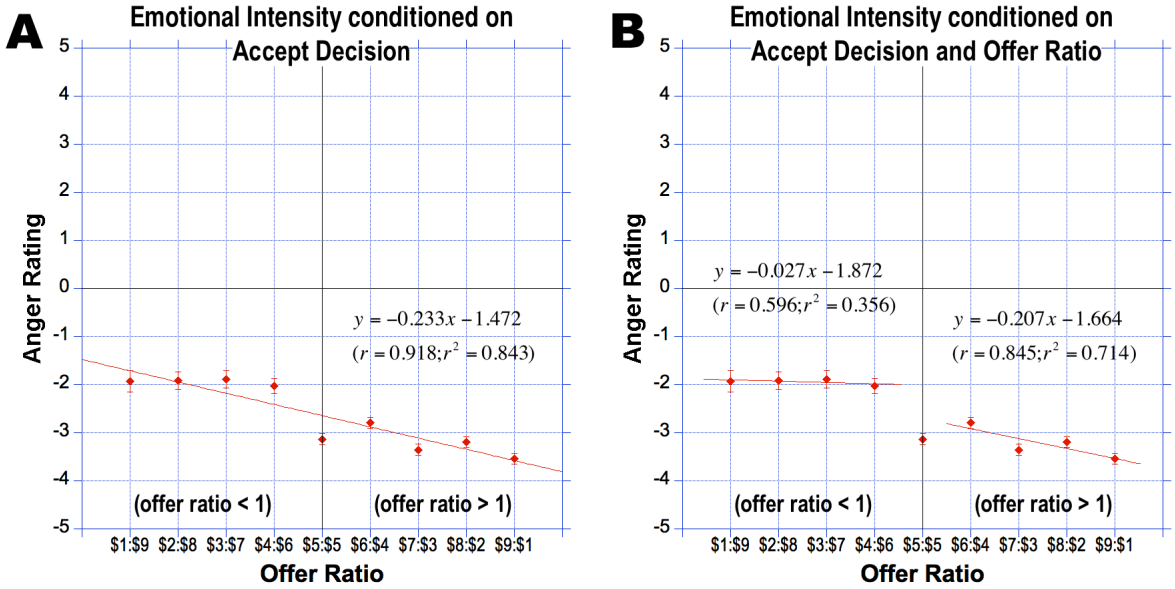


Fig. 2 Average emotional angry response to the monetary offer conditioned on acceptance trials only. **a** Curve-fitting to acceptance trials shows a linear inverse proportional relationship. **b** Curve-fitting conditioned on hyper-fair (generous) and unfair (stingy) trials shows different proportionality relationship, represented by different slopes. It shows anger is inversely proportional to monetary gain (or proportional to loss), for both hyper-fair and unfair trials, even though there is a discrete change in emotional sensitivity from unfair to hyper-fair trials. The subjects reported not-angry in both hyper-fair and unfair trials when they get either money and/or fairness

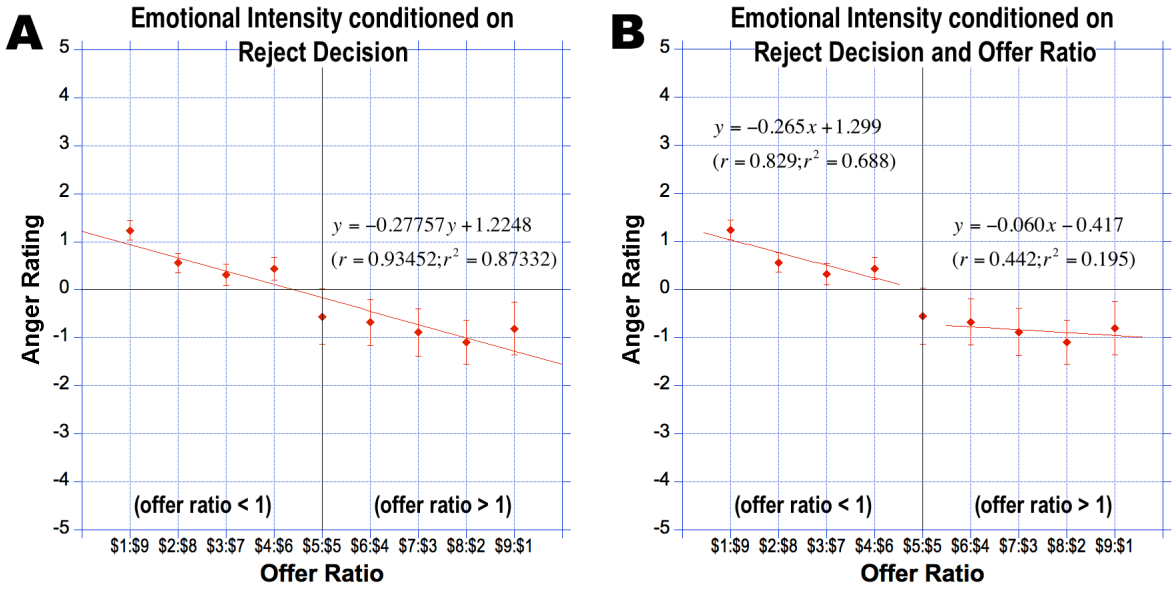


Fig. 3 Average emotional angry response to the monetary offer conditioned on rejection trials only. **a** Curve-fitting to rejection trials shows a similar linear inverse proportional relationship with much higher emotional intensity than Fig. 2, shifting from not-angry to angry. **b** Curve-fitting conditioned on unfair (stingy) offers and hyper-fair (generous) offers shows different proportionality relationships than those in the acceptance trials. The subjects reported angry to unfair offers (losing

both money and fairness), but not-angry to hyper-fair offers (losing money only but not fairness) when they decided to reject the offers (Note that the standard error bars are large in the right-half of the figure because of the smaller sample size, i.e., a small number of subjects rejected hyper-fair offer)

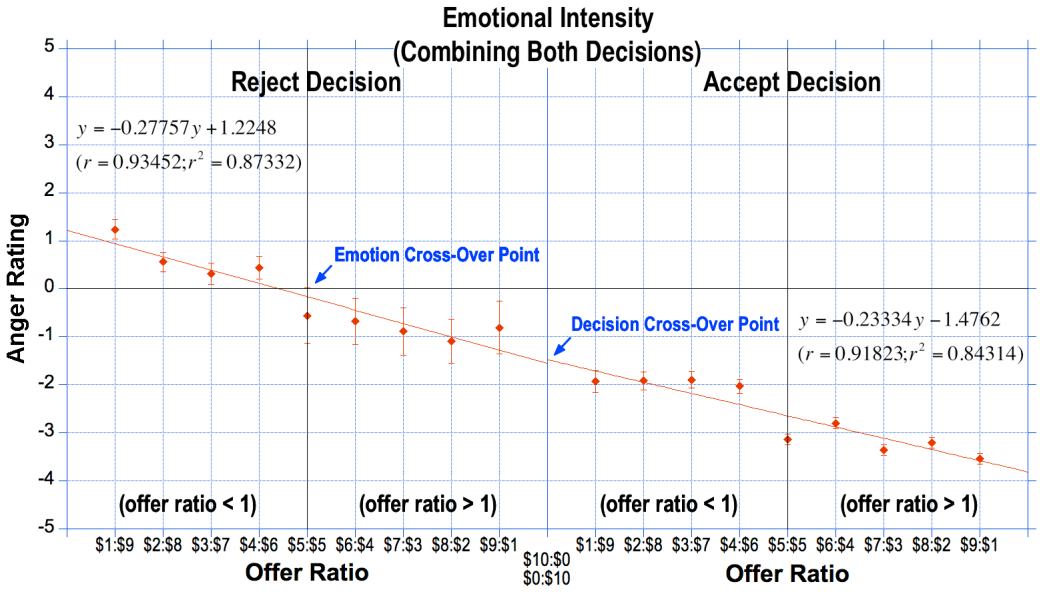


Fig. 4 Graph re-plotted by combining Figs. 2a with 3a to show the continuum in emotional intensity from rejection to acceptance decision

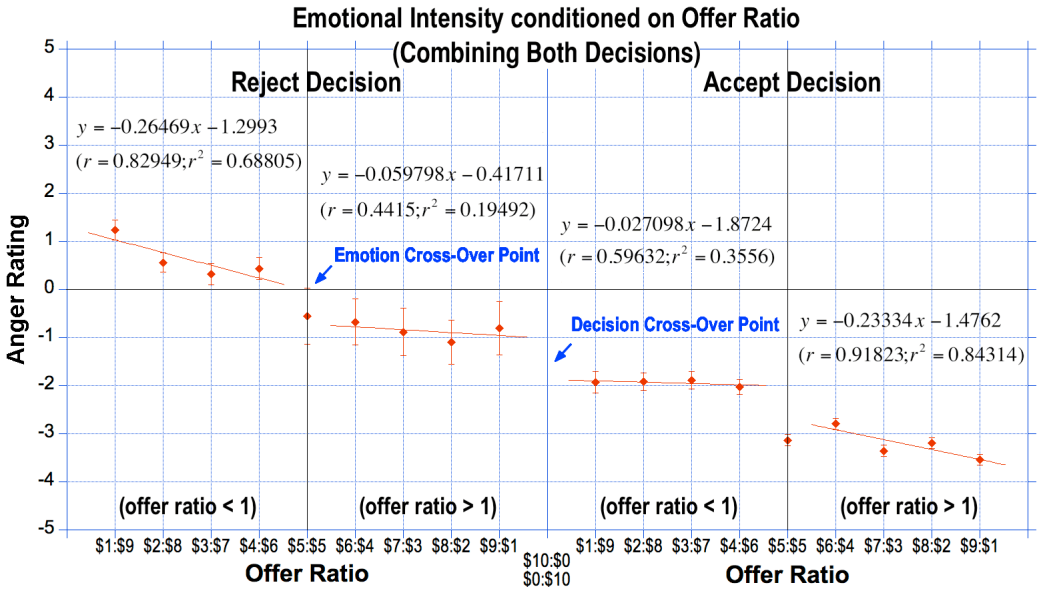


Fig. 5 Graph re-plotted by combining Figs. 2b with 3b to show the step functions in emotional intensity from rejection to acceptance decision