Final version of this paper appears at International Conference on Affective Computing and Intelligent Interaction, 2009

Assessing the validity of appraisal-based models of emotion

Jonathan Gratch, Stacy Marsella, Ning Wang, Brooke Stankovic
University of Southern California
13274 Fiji Way, Marina del Rey, CA 90405

Abstract

We describe an empirical study comparing the accuracy of competing computational models of emotion in predicting human emotional responses in naturalistic emotion-eliciting situations. The results find clear differences in models' ability to forecast human emotional responses, and provide guidance on how to develop more accurate models of human emotion.

1. Introduction

In recent years, research in emotion has expanded over a wide range of disciplines, in the process revising our understanding of the role emotion plays in human behavior. Increasingly, modern theories of emotion posit that emotion plays a functional, often beneficial role, in cognition and behavior [1, 2]. Research in the neurosciences has helped to identify how emotion plays a central role in decision-making. Work in social psychology argues that emotion is critical to human social interaction. In economics, the study of emotions has revised theories of economic decision-making [3].

Coupled to this growing body of research, there has been extensive work in computational models of human emotion. Emotion models have been proposed to improve the modeling of users in order to facilitate improved human-computer interaction [4] and create more effective intelligent tutoring systems [5]. Emotion models have also been posited as a way to improve the performance of intelligent systems or robots by making them more robust and reactive [6]. Emotion models have become a core component in embodied agent research, where emotion is used to create more life-like, expressive virtual characters for a variety of applications. In particular, virtual humans are being employed for entertainment [7] as well as for virtual reality based social skills training [8], where a learner practices difficult social interactions with life-like virtual characters. The development and simulation of computational models of emotion have also been proposed as a basic research methodology for exploring the dynamic properties of human cognition and emotion [2].

Although there has been extensive work in computational models of emotion, surprisingly little work has been done in validating these models, with a few exceptions [9, 10]. Accordingly, researchers and developers alike are faced with a confusing array of models with little guidance on which are more appropriate for their work. We can define several kinds of validation depending on the application of the model. One might assess if they improve a particular application in which they are used: e.g., in human-computer interaction, does a model of the user's emotions create a more efficient or satisfying user experience? Similarly, one might ask if an emotion model improves learning in a tutoring system (as in [5]) or interactive entertainment systems more emotionally-evocative. In this article, we focus on the criteria of *behavioral fidelity*: assessing the consistency between predictions made by the model and the behavior of human subjects in naturalistic emotion-eliciting situations. To assess behavioral fidelity, we can, for example, assess both the antecedents and consequences of emotions, including whether the model correctly predicts what emotions a person will have in a particular situation, the intensity of those emotions, the temporal evolution of those emotions as well as the impact of emotions on cognitive processes and behavior.

This article describes the results of a rigorous empirical study contrasting the behavioral fidelity of alternative computational models of appraisal theory. We survey computational models of emotion proposed in the literature and identify contradictory assumptions adopted by these methods concerning how they derive emotional responses from a situational appraisal. We contrast these assumptions with behavior of human subjects playing a competitive board game, using monetary gains and losses to induce emotion. We indexed subject's appraisals and emotional state at key points throughout a game, revealing a coherent pattern in the dynamic relationship between these factors. The results provide clear guidance for the appraisals to seek to faithfully model human emotional behavior.

2. Computational Appraisal Models

Appraisal theory argues that emotions should change as a function of how a situation is appraised (good vs. bad; likely vs. unlikely; controllable vs. uncontrollable; etc.). For example, imagine the situation in **Figure 1**a where Mary is playing a game with probability of winning, p, positive utility associated with winning (U_{WIN}), and negative utility associated with losing (U_{LOSS}). Mary's emotions, according to appraisal theory, are determined by her subjective sense of the probability and utility of these outcomes, and possibly other appraisal factors. Multiple emotions are possible and the intensity of each emotional response depends on the current value of appraisal variables. For example, in **Figure 1**b, Mary is playing a game where she'll receive $5 for winning and lose $1 for losing. She believes she is winning and feels predominantly hopeful.

Appraisal theory provides, at best, a high-level specification for a computational model of emotion, forcing modelers to adopt representational and process assumptions to create a working system. Computational appraisal models make differing choices including 1) how to represent situations, 2) which appraisals to compute, 3) how appraisals relate to the intensity of emotion, 4) how different emotions combine into an overall emotional state 5) how this emotional state changes over time and as the situation evolves, and 6) how emotions alter subsequent actions and the interpretation of the situation. Alternative computational models typically differ on multiple dimensions, making comparative studies tricky, at best.

This paper reports the first in a series of studies evaluating the validity of alternative design choices proposed within the computational appraisal literature. Here, we describe and evaluate the behavioral fidelity of competing proposals for determining how appraisals relate to the intensity of an emotion.

2.1 Intensity Models

Most computational appraisal models provide specific equations that predict how appraisal variables impact the intensity of an emotional response, though these equations are rarely consistent across models and in some cases make contradictory predictions. For example, El Nasr's FLAME [11] proposes the amount of joy resulting from an event is an additive function of probability and utility of goal attainment.[1] In contrast, Neal Reilly's EM [7] calculates joy by multiplying the utility of a goal by the change in probability of goal attainment that results from an emotion-eliciting event.

We surveyed several computational appraisal models and grouped their intensity equations into a small number of generalized *intensity models*. These intensity models highlight alternative approaches for synthesizing the contribution of two central appraisal dimensions: probability and utility (often referred to as desirability, pleasantness or goal conduciveness within the appraisal literature).[2] This is not to say that other appraisal dimensions are unimportant -- for example, Elliott's AR model [12] mentions 22 intensity moderating variables (although how they are utilized is not explicitly described) and more recently Marinier [13] proposed an intensity equation involving probability, discrepancy from expectation, suddenness, unpredictability, goal relevance, intrinsic pleasantness, conduciveness, control and power. However, several experimental studies suggest probability and utility explain a good portion of the variance in affective responses and all of the computational approaches we have surveyed agree on the importance of these variables. Thus, we argue the relationship between these fundamental factors must be settled before it is profitable to bring in additional complicating factors.

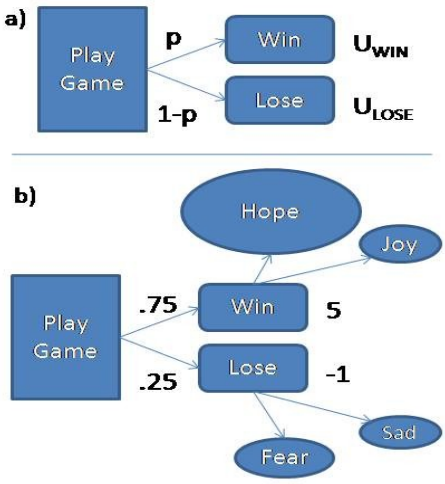

Figure 1: The representation of an emotional situation

TABLE 1	Hope	Joy	Fear	Sadness
Expectation-change	PEACTIDM	ParleE, EM PEACTIDM	PEACTIDM	ParleE, EM PEACTIDM
Expected Utility	EMA, ParleE, FearNot! EM, BDTE		EMA, ParleE, FearNot!, EM BDTE	
Threshold		EMA, FearNot! BDTE		EMA, FearNot! BDTE
Additive	Cathexis. FLAME	Cathexis, FLAME	Cathexis, FLAME	Cathexis, FLAME
Hybrid	Price et al85	Price et al85	Price et al85	Price et al85

Our proposed intensity models contain free parameters to emphasize our focus on evaluating the general form of the intensity relationship, not the specific coefficients proposed by specific authors. For example, Price et al [14] define a specific intensity relationship:

$$\text{Intensity} = 1.7 \times (\text{Utility} \times \text{Probability})^{0.5} - 0.7 \times \text{Utility}$$

Note that this equation represents the intensity of an emotional response as nonlinear *power function* of probability and utility (i.e., raising these factors to some power). Power functions are common in studies of emotion and decision-making as people's subjective perception of probability and utility rarely follow simple linear relationships (see [14-16]). Indeed, some of the models we contrast use power functions and the rest can be interpreted as degenerate power functions (with an exponent of 1.0). In defining generalized intensity models, we treat the intensity of an emotional response as some power function of probability and utility. We further included a set of "typical" parameter settings (i.e., exponents and coefficients) that correspond to the typical settings found in models of this class.

Table 1 summarizes the intensity models we will explore in this paper and some of emotion systems that adopt these models. Note that systems often used different intensity models for different emotions and some

[1] In papers, it is ambiguous if FLAME uses an additive or multiplicative model. In personal communication with Prof. El Nasr we clarified it uses an additive model.

[2] Note that some models distinguish between the utility of winning and the dis-utility of losing. We revisit this in the discussion.

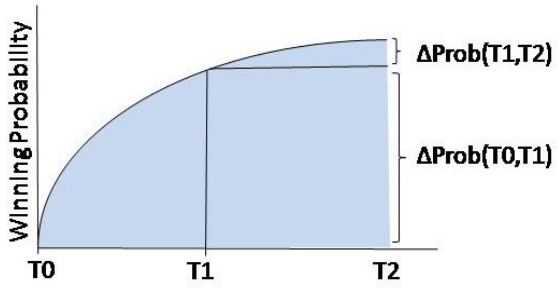

Figure 2: Intensity predicted by expectation-change model

models are never used for certain emotions (for example, no system we reviewed suggests using expected utility as a way of modeling the intensity of Joy.

Expected Utility Models: Perhaps the most common approach is to base emotional intensity of an event on some variant of expected utility (e.g., see EM [7], EMA [17], FearNot! [18], PARLE [19], BDET [20]). For example, when calculating the amount of hope to associate with a future action, EMA multiplies the utility of any goal achieved by the action by the probability that the action will achieve this goal. The generalized form of the expected utility model is of the form:

- $HOPE_{EU} = a \times U^p \times P^q + b$ if $P < 1.0$, else $= b$
- $JOY_{EU} = a \times U^p \times P^q + b$
- $FEAR_{EU} = a \times U^p \times (1 - P)^q + b$ if $P > 0$, else $= b$
- $SAD_{EU} = a \times U^p \times (1 - P)^q + b$

where the typical parameters would be a=p=q=1, b=0

Expectation-change Models: Several models (e.g., EM, ParleE [19], PEACTIDM [13]) tie intensity of emotional response to changes in the likelihood of an event. For example, Neil Reilly argues that expected stimuli should be less intense than unexpected stimuli [21]. This argues that intensities should be a function of the change in probability of goal attainment (ΔP) that an event produces. For example, EM proposes:

$$JOY_{EM} = U \times \Delta P$$

Similarly, Marinier proposes the intensity of response is proportional to $(1\text{-}OP)(1\text{-}DE)+(OP \times DE) \times U$, where $OP=P_{t-1}$ and $DE=abs(\Delta P)$.

To cover the various approaches that utilize this approach, we propose the following generalized intensity equations (with typical parameter values a=p=q=1, b=0):

- $HOPE_R = a \times U^p \times \Delta P^q + b$ if $\Delta P > 0$; else $= 0$
- $JOY_R = a \times U^p \times \Delta P^q + b$ if $\Delta P > 0$; else $= 0$
- $FEAR_R = a \times U^p \times |\Delta P|^q + b$ if $\Delta P < 0$; else $= 0$
- $SAD_R = a \times U^p \times |\Delta P|^q + b$ if $\Delta P < 0$; else $= 0$

Several studies show empirical support for this model for a special case: where the outcome of an uncertain action becomes revealed. For example, Mellers et al. [16] showed elation and disappointment are greater when the outcome of a gamble was unexpected. However, such studies do not address what happens when a situation unfolds over time.

Threshold Models: Some emotion models have intensity equations based on the probability crossing some threshold. For example, EMA, FearNot! and BDET [20] produce joy only when a desired outcome becomes certain (i.e., P=1.0).

We abstract these approaches by incorporating a threshold into the expected utility model (where typical parameter settings would be a=p=q=1, b=0 and the threshold t varies by emotion type):

- $HOPE_{TH} = a \times U^p \times P^q + b$ if $t \leq P < 1.0$; else $= 0$ ($t \approx 0.5$)
- $JOY_{TH} = a \times U^p \times P^q + b$ if $P \geq t$; else $= 0$ ($t \approx 1.0$)
- $FEAR_{TH} = a \times U^p \times (1 - P)^q + b$ if $0 < P < t$; else$=0$ ($t \approx 0.5$)
- $SAD_{TH} = a \times U^p \times (1 - P)^q + b$ if $P \leq t$; else $= 0$ ($t \approx 0.0$)

Additive Models: Some emotion models (FLAME [11], Cathexis [22]) have intensity equations based on the sum of probability and utility. For example, FLAME propose a set of rules loosely inspired by the work of Price et al [14]:

- $Hope = 1.7 \times P^{0.5} + -0.7 \times U$
- $Fear = 2 \times (1\text{-}P)^2 - U$

And Cathexis derives an emotional intensity as a linear combination of several appraisal factors.

Such additive models can be abstractly characterized by the following set of equations (with preferred parameter values being a=p=q=b=1):

- $HOPE_{ADD} = a \times U^p + b \times P^q$
- $JOY_{ADD} = a \times U^p + b \times P^q$
- $FEAR_{ADD} = a \times U^p + b \times (1 - P)^q$
- $SAD_{ADD} = a \times U^p + b \times (1 - P)^q$

Hybrid models: Price et al. [14] propose a hybrid model with both a multiplicative and additive relationship between probability and utility. We include this here because it has some empirical support and inspired FLAME's intensity equations.

2.2 Contradictory Predictions

So far we have demonstrated that different appraisal approaches use different equations to derive the intensity of emotional response, but are these differences significant? In fact, it is easy to show that they make contradictory predictions.

Figure 3 illustrates the predicted intensity of emotion of different models as a function of the probability of goal attainment and the utility of the goal (for three hypothesized utility values – 1, 2 and 3 – and assuming a=p=q=1.0 and b=0.0. The expected utility model exhibits a relationship referred to as a *linear fan pattern* [23]: utility increases monotonically with the probability of goal achievement, but at a different rate depending on the utility of the goal. In contrast, the additive model exhibits a parallel pattern and the threshold model exhibits a sharp transition in emotion intensity as goal attainment becomes certain.

The expectation-change model is more complex as the intensity of emotion depends on the change in probability over time, but again, it is easy to illustrate that models make different predictions. To illustrate, let us return to our example of Mary playing a game. Imagine that Mary is winning and that **Figure 2** describes how her probability of winning changes over time. As the probability

change from T0 to T1 is greater than the change from T1 till T2, a expectation-change approach such as PEAC-TIDM would predict that the intensity of hope would decrease over time. An expected utility approach such as EMA would predict that hope would increase over time. Thus, two different models would make opposite predictions for the same situation.

3. Experiment

To evaluate different intensity models we manipulate subjects' perceptions of the probability and utility of goal attainment and assess their emotional state. We compare the predictions of intensity models with subjects' responses when playing a competitive board game called Battleship™ by the Milton Bradley Company. In the standard game, players secretly place ships on a small grid, then take turns shooting at squares in the grid in an attempt to sink their opponent's ships.

To induce emotions, subjects play for money (they can win or lose up to $10 US). To induce positive and negative emotions we alter perceptions of winning likelihood (within and between subjects) via the sequence of hits and misses in the game, and perceptions about winning/losing importance (between subject) by framing the game as an opportunity to win or to lose money. Game play is altered via a confederate: Although subjects believe they played another subject, in reality they played against a confederate watching their game play through a hidden camera and controls the series of hits and misses.

We also would like to assess how subjects' emotions unfold over time. To explore appraisal dynamics, we use repeated measures to assess how subjects' emotions change within the game. We index subjects' subjective impressions at the game's start, middle and end.

3.1 Method

One-hundred and seven people (41% women, 59% men) were recruited via craigslist.com from the greater Los Angeles area. Subjects were compensated $30 for one hour of their participation. Participants ranged from 18 to 60 (36 average), 4% with some high school education, 39% with some college education, 45% with college diploma, and 12% with graduate degree.

3.1.1 Design

The study is a 2x2 between-subjects design. The two independent variables are framing and game play.

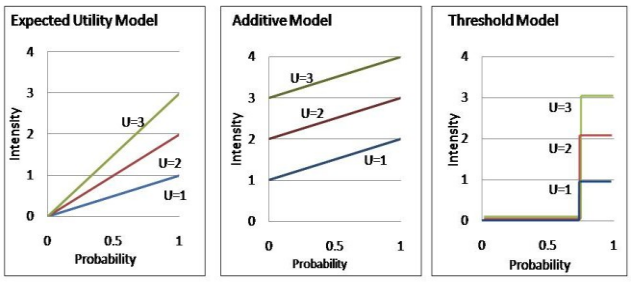

Figure 3: Predicted intensities of emotion as a function of probability and utility of goal attainment

Framing. There are two conditions for the framing: positive (n=48) and negative incentive (n=59). In the positive incentive condition, participants are recruited using posters saying they will be paid $20. Upon arriving at the lab, they are then informed that they can win up to additional $10 if they win the game. In the negative incentive condition, the recruitment poster says the compensation is $30. When arrived at the lab, the experimenter informs the participants that they can lose up to $10 if they lose the game. All participants are paid $30 in the end regardless of framing and game result.

Game play. There are two conditions for the game play: win (n=53) and lose game (n=54). In the winning game play condition, participant wins the game. In the losing game play condition, participant loses the game.

3.1.2 Procedure

Participant and the confederate enter the laboratory and are told they are participating in a study on games. After they read the consent form, the experimenter explained to the participants in the positive incentive condition that the winning player can win up to additional $10. The participants in the negative incentive condition were told that the losing player can lose up to $10.

The confederate and the participant view a PowerPoint presentation about rules of Battleship™ and experimental procedures. They then fill out a pre-test questionnaire. Battleship game began after completion of the questionnaire. Experimenter leaves the room.

Game play is divided into three stages. The start of the game (*T0*), the point when one player will likely win (with 75% probability),[3] and when the game has been won by one of the players (*T2*). At each point the participant fills out an appraisal and emotion questionnaire.

Finally, participants were debriefed individually and probed for suspicion using the protocol from Aronson, et al. [24]. No participants indicated that they believed their opponent was a confederate in the study. All subjects were allowed to retain the addition $10.

3.1.3 Equipment

The participant and confederate face each other across a desk separated by a white board that blocks their view. The game and a desktop computer is in front of each. The participant fills out the questionnaires on the computer. A hidden wireless camera is placed on the ceiling to record participant's moves on the Battleship board. The camera video is sent to the confederate's computer.

3.1.4 Measures[4]

Demographic/Dispositional information: At the beginning of the experiment we ask participants demographic information, board game and Battleship experience, and

[3] A pilot study found this to be when the participant has sunk 3 of the confederates 4 ships (win condition) or lost 3 of 4 of their own ships (lose condition).

[4] Several measures are included for completeness but not discussed as they apply to questions in a companion article [25].

a measure of tendency to be cooperative, individualistic or competitive. [26].

Several items are repeatedly measured at time points *T0* (start), *T1* (middle), and *T2* (end):

Emotions. Intensity of self-reported emotions were elicited with a visual analog scale ranging from 0 to 100. We assessed fear, joy, sadness, anger and hope.

Appraisal and Coping Scale. We developed a 12-item appraisal scale to measure participant's perceptions of winning utility and likelihood, ability to control the outcome, effort devoted to winning, as well as several measures related to importance and likelihood that the game was played fairly.

All scales are presented as an analog scale that ranges from zero (minimum value/intensity) to 100 (maximum value/intensity).

3.2 Results

Data from six sessions were excluded due to incomplete questionnaire or because experiment procedure deviated from the standard procedure. Data from 101 participants were included in the analysis, 48 from the losing condition and 53 from the winning condition.

3.2.1 Manipulation check

Our manipulation of subjective sense of winning was successful. Subjects perceived they have an approximately even chance of winning at the start of the game. Perceptions of winning increased in the wining conditions ($p<0.001$) and decreased in the losing condition ($p<0.001$). Perceived probability changed approximately linearly across the stages of the game: Pr(Losing)=0.27; Pr(Start)=0.55; Pr(Wining)=0.76.

Our manipulation of incentive was unsuccessful – subjects' responses were largely indistinguishable.[5] However, both positive and negative emotions were successfully elicited and we collapse across the incentive conditions in all subsequent analysis.

3.2.2 Intensity results

Figure 4 summarize the reported intensity of emotion across conditions (Win vs. Lose) and at the different points sampled during the game (T0=start; T1=middle; T2=end). There are significant differences in emotional state as a function of both condition and time. Positive emotions are significantly more intense than negative emotions. There are also significant interactions by condition and by time.

We performed two analyses on the intensity results. A qualitative analysis examined if there were significant changes in intensity that mach to the predictions of competing intensity models. Second, a more detailed quantitative analysis assessed which intensity model best fits the intensity data.

[5] MANOVA showed no interaction between framing and game play on any independent variables (hope, fear, joy, sadness) except for a small significant interaction with fear when participants are losing/winning.

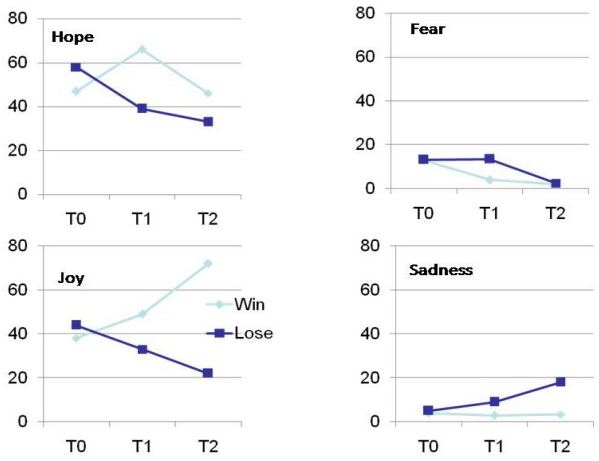

Figure 4: Self-reported emotional intensity

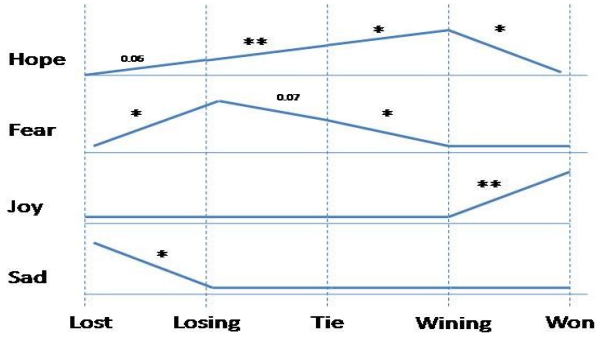

Figure 5: Qualitative emotion intensity results

Table 2	Joy	Sad	Hope	Fear
Expected Utility	0.801	0.829	0.932	0.922
Expectation-change	∅	∅	∅	∅
Threshold	∅	∅	∅	∅
Additive	0.766	0.685	0.596	0.295
FLAME	∅	∅	∅	∅
Price & Burrell	∅	∅	∅	∅

In qualitative terms, the models make differing predictions about how emotion intensity should change at different states of the game. For example, a threshold model for joy predicts no significant changes in joy intensity until the game was won; the expected utility and additive models predict joy will increase monotonically as winning probability increases; and a expectation-change model predicts that joy will remain constant throughout as ΔProbability is approximately constant across stages of the game.

Figure 5 illustrates the significant qualitative changes in emotion intensities as a function of the stage of the game (collapsing across the win and loss condition). Asterisks indicate the significance level of intensity changes ('*' indicates $p\leq0.05$; '**' indicates $p\leq0.01$). These results lend some support for a threshold model for joy and sadness and both the expected utility and additive models for hope and fear.

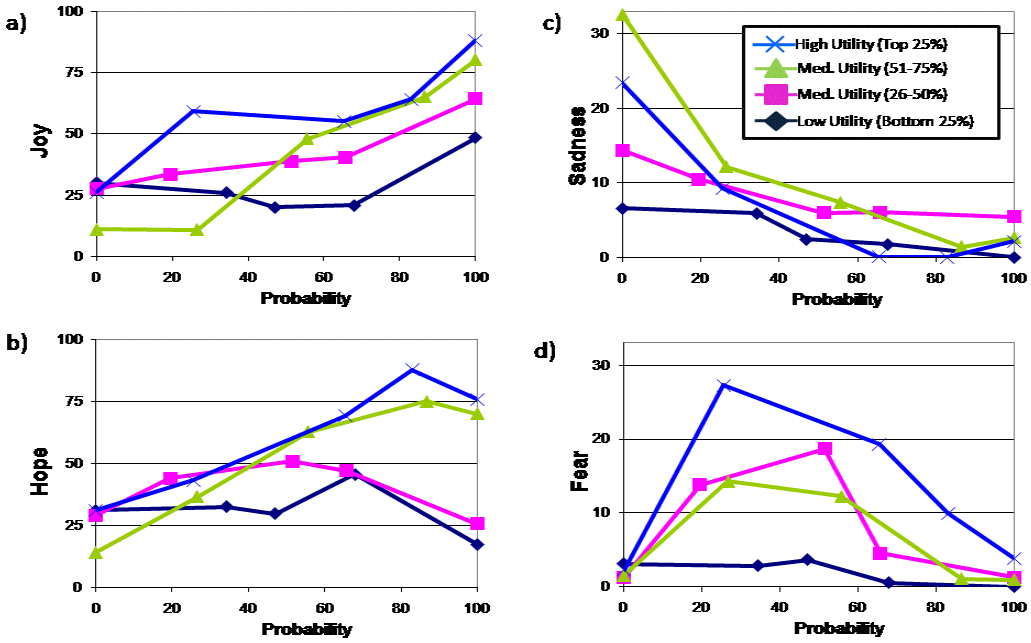

Figure 6: Self-reported emotional intensity as a function of probability and utility of goal attainment

Second, we applied a detailed quantitative analysis examining the fit of each model according to the method of Anderson [27]. In this method, subjects are grouped by their initial desire to win: we performed a quartile split, dividing subjects into four groups as a function of the utility they assigned to winning at T0 (e.g., the quarter of subjects that assigned the most utility to winning at T0 were combined into a *high-utility group*). Figure 6 shows the emotional intensity of these different groups as a function of winning likelihood. Each line illustrates the average emotional intensity response of subjects in each group at each time point (i.e., lost, losing, start, wining, won).[6] For example, Figure 6d illustrates that after losing or wining the game, subjects reported they felt no fear, but that fear in the middle of the game increased as a function of how much utility they assigned to winning (high-utility subjects had an average fear response of 25.6 when losing, whereas low utility subjects had an average fear response of only 3.6).

To contrast the quality of different intensity models we performed nonlinear regression over the graphs in Figure 6, comparing the fit of different intensity models proposed above. Nonlinear regression requires initial starting values for the parameters to be fit and we use the default values listed above. Table 2 summarizes the results (N=101). The table lists r^2 values (a measure of goodness-of-fit that ranges from zero to one). Any value above 0.7 is considered a good fit. The symbol ∅ indicates that the function had a worse fit than simply using the mean of the variable predicted (i.e., the model is bad). This analysis lends support for the expected utility model for all emotions, with a particularly strong fit for the prospective emotions (i.e., hope and fear).

The parameters for the expected utility model that yielded the best fit to the data are:

- Joy = $1.41 \times U^{0.83} \times P^{1.54} + 2.37$
- Sad = $0.60 \times U^{0.82} \times (1\text{-}P)^{3.06} + 2.32$
- Hope = $0.02 \times U^{1.45} \times P^{1.0} + 1.45$ where $P < 1.0$
- Fear = $0.79 \times U^{0.98} \times (1\text{-}P)^{1.21} + 30.38$ where $P > 0.0$

3.3 Discussion

The results show clear differences in the behavioral fidelity of different intensity models.

The expected utility model provides an excellent fit for modeling the intensity of the prospect-based emotions (i.e. hope and fear). Most appraisal models assume that the intensity of hope/fear drops to zero at this point. This assumption held for fear but not hope. Hope dropped when subjects won the game but not as much as predicted. This might be explained by arguing that subjects are hopeful of receiving their payment (a factor not directly teased apart in this study).

The fit for both outcome-based emotions (i.e. joy and sadness) is also well-approximated by an expected utility model. Unlike hope and fear, however, the exponents identified by regression differ considerably from 1.0. Joy and sadness grow superlinearly with probability, particularly for sadness which grows with the cube of probability. As this exponent increases the shape of the curve converges to a threshold model, suggesting that joy and sadness lie somewhere between a threshold and expected utility model. Indeed, the qualitative analysis showed that the only significant differences were consistent with the threshold model of joy and sadness.

Other intensity models did not fare well. The expectation-change model was worse than a degenerate model

[6] Note that Figure 7 collapses results across conditions. Data on the left half of the graph are from the Lose condition. Data on the right half are from the win condition. We averaged all subjects to obtain the perceived probability at T0.

that ignores probability and utility. The additive model performed better than the expectation-change model but was dominated by expected utility. Specific versions of the additive model, e.g., the model used in FLAME and Price et al., were a poor fit.

3.3.1 Framing effects

There are several limitations and qualifications to the study. One issue is the current study correlated intensity of emotion with self-reported utility of winning, however people may assign different utility to gains vs. loses [15]. Indeed, we included a positive vs. negative incentive condition because we expected subjective utility of winning to differ if the game was framed as a financial gain or loss. The failure of this manipulation suggests that money wasn't the primary motivator for subjects.

Regardless of their motives, the fact that subjects reported more intense positive than negative emotions suggests asymmetries in the subjective utility of losing and wining. Our generalized intensity models are robust to asymmetry as they allow a linear scaling of utility across different emotions (e.g., Joy $\approx 1.4\times U$ and Sad $\approx 0.6\times U$). However, more precise fits could be obtained if we elicited both utility and disutility measures. This may be important for disentangling why our positive/negative incentive manipulation failed to show an effect.

3.3.2 Granularity of Representation

One complication is that situations can be represented at varying levels of detail. For example, we assumed that the game is represented by a single goal (win) and a single abstract action (play-game) and that subjects' self-reported hope represents their emotions associated with the goal of winning. However, there are many levels of detail one could choose to represent this situation. Indeed, several appraisal models allow hierarchical representations of situations. For example, one could model the situation at the level of ships: i.e., players have subgoals associated with sinking each of their opponent's ships and maintaining their own. One could also model the situation at the level of turns – i.e., each move has a probability of hitting or missing and may elicit an emotional response. If one allows hierarchical representations of situations, than self-reported emotions such as hope may well represent a summary of multiple emotions: i.e., the hope I will win as well as the hope that I will sink the opponent's ship in the next turn.

If they occur, these "subgoal emotions" would likely skew the regression analysis as they would be correlated with perceptions of winning. For example, in the Win condition, subjects sink a series of ships, leading to intermediate instances of joy, while simultaneously increasing the perceived likelihood of winning. One consequence of this is that even if the underlying model for joy and sadness was a threshold model, our regression analysis would interpret this as an expected utility model as subgoal emotions are credited to the top-level goal of winning. In fact, this is exactly the behavior we would expect from EMA (which uses a threshold model for joy and sadness) if we represented these subgoals explicitly.

3.3.3 Unfolding situations

The expectation-change model fared poorly in this study but it has received support in other studies and these differences need to be explained. For example, Mellers et al. [28] showed that ΔProbability was a good predictor of emotion intensity for emotions such as elation and disappointment and Reisenzein showed it is a key moderator of surprise [20].

Of course we studied different emotions, but more significantly, battleship is a different type of situation than those explored in those studies. Mellers and Reisenzein studied "single-step" situations where the outcome of an uncertain event is instantaneously revealed. Specifically, they studied simple bargains (e.g., you have a 25% chance of winning $10 and a 75% chance of losing $5). In contrast, battleship unfolds over time through a series of steps. What our results suggest is that the expectation-change model may be a poor fit for these sort of unfolding situations.

Unfolding situations are problematic for the expectation-change model in a more fundamental sense as the change in probability is defined with respect to a reference point. When a situation gradually unfolds over time, this reference point is ill-defined. In battleship, should it be set at the start of the game? the previous time a ship was sunk? at the last move? In the results above we used the change in probability from the last time point sampled (i.e., the change from T0 to T1 or from T1 to T2), although one could make other choices. If one chooses the initial point (T1), then the model reverts to an expected utility model. If one uses the previous move in the game, then the change of probability of winning is essentially zero.

3.3.4 Other appraisal factors

The current study focused on appraisals of utility and probability, but other factors claimed to be important in determining the intensity of emotional response. One way to address this is to explore correlations between other factors and emotional response. The current study did not systematically explore this question.

3.3.5 Self report

A final point to emphasize is that, although this study is a considerable advance over prior evaluations that utilized self-reported emotional responses from imaginary situations, it still relies on self report to elicit subjects' appraisal and emotional responses. Such work is always open to the criticism that results are subject to impression-management biases or that they represent subjects' beliefs about emotion rather than the emotion itself. Future work is focusing on the buttressing these findings with a variety of behavior measures.

3.4 Implications for appraisal models

These results cast doubt on the behavioral fidelity of several computational appraisal approaches. Revisiting Table 1, the results are inconsistent with the models adopted by PEACTIDM, FLAME and Cathexis. The results are partially inconsistent with EM and ParleE. Of the methods we surveyed, only EMA and FearNot! are consistent with the results of our study.

It is important to note that not all appraisal approaches aim for behavioral fidelity, and the fact that a model is inconsistent with human data doesn't necessarily mean it will be ineffective. EM, for example, is intended for interactive entertainment and Neal Reilly disavows any claims concerning the fidelity of his proposed intensity models [21]. However, in the absence of evidence that unrealistic intensity equations enhance user perceptions, it seems preferable to incorporate intensity models with clear behavioral fidelity.

4. Conclusions

In this paper, we describe the results of an empirical study contrasting methods for predicting the intensity of an emotional response from appraisals. We identified and reviewed several approaches that have been proposed in the literature and contrasted them with behavior of human subjects playing a competitive board game, using monetary gains and losses to induce emotion. Results can be used to help distinguish the validity of competing approaches to modeling emotion.

5. Acknowledgments

We thank Anya Okhmatovskaia and Wendy Treynor for help in study design. Rainer Reisenzein and Scott Neal Reilly provided thoughtful suggestions. This work was sponsored by the U.S. Army Research, Development, and Engineering Command and the Air Force Office of Scientific Research under the grant #FA9550-06-1-0206. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

6. References

1. Keltner, D. and J. Haidt, *Social Functions of Emotions at Four Levels of Anysis.* Cognition and Emotion, 1999. **13**(5): p. 505-521.
2. Gratch, J. and S. Marsella, *The Architectural Role of Emotion in Cognitive Systems*, in *Integrated Models of Cognitive Systems*, W. Gray, Editor. 2007, Oxford.
3. Loewenstein, G. and J.S. Lerner, *The role of affect in decision making.*, in *Handbook of Affective Science*, 2003, Oxford: Oxford University Press. p. 619-642.
4. Picard, R.W., *Affective Computing*. 1997, MIT Press.
5. Conati, C. and H. MacLaren. *Evaluating a probabilistic model of student affect*. in *7th International Conference on Intelligent Tutoring Systems*. 2004. Maceio, Brazil.
6. Scheutz, M. and A. Sloman. *Affect and agent control: experiments with simple affective states*. in *IAT*. 2001: World Scientific Publisher.
7. Neal Reilly, W.S., *Believable Social and Emotional Agents*. 1996, CMU: Pittsburgh, PA.
8. Swartout, W., et al., *Toward Virtual Humans.* AI Magazine, 2006. **27**(1).
9. Mao, W. and J. Gratch. *Evaluating a computational model of social causality and responsibility*. in *AAMAS*. 2006. Hakodate, Japan.
10. Gratch, J. and S. Marsella. *Evaluating a General Model of Emotional Appraisal and Coping*. in *AAAI Symposium on Architectures for modeling emotion: cross-disciplinary foundations*. 2004. Palo Alto, CA.
11. El Nasr, M.S., J. Yen, and T. Ioerger, *FLAME: Fuzzy Logic Adaptive Model of Emotions.* Autonomous Agents and Multi-Agent Systems, 2000. **3**(3): p. 219-257.
12. Elliott, C. and G. Siegle, *Variables influencing the intensity of simulated affective states*, in *AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*. 1993, AAAI: Palo Alto, CA.
13. Marinier, R., *A Computational Unification of Cognitive Control, Emotion, and Learning*, in *Computer Science*. 2008, University of Michigan: Ann Arbor, MI.
14. Price, D.D., J.E. Barrell, and J.J. Barrell, *A quantitative-experiential analysis of human emotions.* Motivation and Emotion, 1985. **9**(1).
15. Kahneman, D. and A. Tversky, *Prospect Theory: An Analysis of Decision under Risk.* Econometrica, 1979. **XLVII**: p. 263-291.
16. Mellers, B.A., et al., *Decision affect theory: Emotional reactions to the outcomes of risky options.* Psychological Science, 1997. **8**(6): p. 423-429.
17. Gratch, J. and S. Marsella, *A domain independent framework for modeling emotion.* Cognitive Systems Research, 2004. **5**(4): p. 269-306.
18. Dias, J. and A. Paiva. *Feeling and Reasoning: a Computational Model for Emotional Agents*. in *12th Portuguese Conference on Artificial Intelligence,* 2005: Springer.
19. Bui, T.D., *Creating emotions and facial expressions for embodied agents*, PhD Thesis, University of Twente.
20. Reisenzein, R., *Emotions as Metarepresentational States of Mind: Naturalizing the Belief-Desire Theory of Emotion.* Journal of Cognitive Systems Research, 2009. **10**(1).
21. Neal Reilly, W.S., *Modeling what happens between emotional antecedents and emotional consequents*, in *Euro. Meeting on Cybernetics and Systems Research*. 2006.
22. Velásquez, J. *When robots weep: emotional memories and decision-making.* in *Fifteenth National Conference on Artificial Intelligence*. 1998. Madison, WI.
23. Anderson, J.R. and C. Lebiere, *The Atomic Components of Thought*. 1998, Mahwah, NJ: Lawrence Erlbaum.
24. Aronson, E., et al., *Methods of Research in Social Psychology*. 2nd Edition ed. 1990, New York: McGraw-Hill.
25. Marsella, S., et al., *Assessing the validity of a computational model of emotional coping*, in *International Conference on Affective Computing and Intelligent Interaction*. 2009, IEEE: Amsterdam.
26. Messick, D.M. and C.G. McClintock, *Motivational Bases of Choice in Experimental Games.* Journal of Experimental Social Psychology, 1968. **4**: p. 1-25.
27. Anderson, N.H., *Information integration approach to emotions and their measurement*, in *Emotion: Theory, research and experience* R. Plutchik and H. Kellerman, Editors. 1989, Academic Press: New York. p. 133-186.
28. Mellers, B.A., et al., *Decision affect theory: How we feel about risky options.* Psychological Science, 1997. **8**: p. 423-429.