

Emotional Perception for Updating Agents’ Beliefs

Bexy Alfonso*, David V. Pynadath†, Margot Lhommet‡ and Stacy Marsella‡

*Departamento de Sistemas Informáticos y Computación, Universitat Politècnica de València

†Institute for Creative Technologies, University of Southern California

‡College of Computing and Information Science, Northeastern University

Abstract—The relative influence of perception and situation in emotional judgments has been extensively debated in psychology. A main issue in this debate concerns how these sources of information are integrated. This work proposes a method able to make probabilistic predictions of appraisals of other agents, using mental models of those agents. From these appraisal predictions, predictions about another agent’s expressions are made, integrated with observations of the other agent’s ambiguous emotional expressions using Bayesian techniques, resulting in updates to the agent’s mental models. Our method is inspired by psychological work on human interpretation of emotional expressions. We demonstrate how these appraisals of others’ emotions and observations of their expressions can be an integral part of an agent capable of Theory of Mind reasoning.

I. INTRODUCTION

Interpreting what others are feeling and thinking is an integral part of our day-to-day lives. The speaker presenting research at a conference scans the people in the audience wondering whether they are entertained, bored or angry. Similarly, the mother of a new born child observes her baby and wonders: is he happy, tired or is he in need of changing? These inferences can draw on several sources of information, including the situation the observed person is in, the observer’s model of the observed’s goals and beliefs and the observed’s behavior such as their facial expressions.

One of the fundamental questions that arises concerns how these sources of information are integrated. A senior professor in the audience seems to be frowning but the speaker’s prior beliefs led him to expect the professor to like the research being presented. Should the speaker interpret that ambiguous frown as negative reaction to the content being presented or is it rather a sign of concentration showing deep interest? Moreover, should the facial expression be a signal that suggests revising his beliefs about the senior professor’s attitudes about the research?

These questions raise fundamental issues that have been debated extensively in psychology. For example, Ekman [1], Tomkins [2], [3], and Izard [4], in the case of expressions, argue that that some facial expressions signal specific “basic emotions”. Those expressions become then, according to this view, a biological adaption, universally recognized, in order to express those basic emotions. In contrast to this “facial dominance” view, Carroll and Russell [5] argue for a “limited situational dominance”. Carroll and Russell argue that situation is fundamental when judging emotions from expression, and that it can even dominate if it is consistent with the expression in terms of arousal and dominance.

We are interested in this issue from the perspective of constructing agents that can integrate bottom-up perception of another agent’s emotional behavior and top-down predictions based on models of that other agent’s beliefs, goals and intentions and use that integration to revise either or both the perception and beliefs. This raises a range of challenges including how to computationally model how these top-down and bottom-up influences are integrated, how the discrepancies between sources are resolved, how to determine which source of information dominates and how the integration influences ambiguous perceptions and beliefs about the other agent.

For our work, we start by drawing on the idea of reverse, or reverse engineered, appraisal [6], [7]. Appraisal theories typically argue that emotion arises from a process of a subjective assessment of the relation between the situation and a person’s concerns. Reverse appraisal essentially inverts this function, trying to make inferences about a person’s personality, goals or beliefs based on their emotional expressions/reactions.

However, how we model the phenomena breaks from the ideas of reverse appraisal in several ways. First, instead of inverting appraisal we argue that the appraisal is operating top-down in the forward direction making (probabilistic) predictions about other agents’ reactions and these predictions are then resolved against bottom-up (probabilistic) perceptions of emotional expressions. The result may alter the perception, the beliefs about the other agent or both. The approach is in line with views that argue the brain is a prediction device [8]. Second, having eliminated reverse appraisal we then proceed to eliminate appraisal, taking an approach in line with Si *et al.* [9] whereby appraisal is a byproduct of social agent’s theory of mind (ToM) processes involved in decision-making and belief revision, in other words that appraisal is an “integral part of the cognitive processes that a social agent must perform to maintain its beliefs about others and to inform its decision-making” [10]. Finally, we seek to model not only perception of emotion expressions role in updating beliefs, but also how the situation and beliefs about the other inform/bias the interpretation of those perceptions.

In this paper, specifically we make a review of some of the approaches that have inspired this work (Section II). In Section III a method is proposed consistent with the top-down operation of appraisals previously described. Subsection III-A describes PsychSim as the supporting tool and modeling of the world and Subsection III-B describes the general steps followed. Section IV describes the implementation of an illustrative example and finally a discussion is offered in

II. SUPPORTING APPROACHES

The way an individual behaves in situations helps another observer to make inferences about that individual [11], [12]. Specifically the perception of emotional expressions is useful to know about the beliefs, goals or character of the observed individual [6]. For example in [13] it is analyzed how this process occurs in negotiation. The effect of one negotiator’s emotions (expressions) on the decisions of other negotiators are analyzed. Specifically, the authors found that when anger is perceived in the opponent, individuals tend to have lower demands and larger concessions, while when perceiving happiness, demands are higher and concessions are smaller. Observations of the emotional feedback is used to modify the belief of what will be the next course of actions of the opponent. In similar experiments [7], C. M. de Melo *et al.* found that the human opponents that perceived guilt in the virtual characters tended to concede more. This result contrasted with previous findings [14], where guilt meant apology and willingness to make amends and hence people tended to concede less. The difference between both results lies in the particularities of the situation when displaying emotions in the judged individual. This brought to the table the importance of the context in the interpretation of emotions.

Closely related is the work by Hareli and Hess [6]. In this work it is argued that it is possible to make inferences about the personality of other people using their emotional reaction through a mechanism of “reverse engineering” that starts from a person’s emotional reaction to derive appraisal information. In this study it was observed for example, that somebody that reacts with anger to an event is perceived as more aggressive than one that reacts with sadness or with a smile. Further, an individual that reacted with anger or with a smile was judged to be more self-confident than the one who reacted with sadness. It was also found that it is the appraisals that mediated the effects, in effect it was the appraisals, not the emotions, that lead to the personality judgments.

From the above mentioned works it is important to consider two different situations. If an individual A is judging an individual B starting from B ’s emotions in reaction to an event e , A may either assess what were B ’s appraisals (and hence B ’s beliefs, goals or personality) by doing a “reverse engineering” (Figure 1(a)), or A can infer, by using its beliefs about B what would be B ’s appraisals (and emotions) given e (Figure 1(b)), and consequently A can either update its beliefs about B based on the emotions perceived and/or modify the emotional perception.

The questions related to the extent to which either the situation (e.g. the event e , A ’s beliefs about B) or the perception (B ’s emotional expression) influence A ’s judgments of facial expressions has been addressed by authors like H. G. Wallbott [15], J. M. Carroll and J. A. Russell [5]. In [15] the author argues for separated judgments of facial expression and context at first instance. Then, if both observations are concordant in relation to the emotions they may suggest, there

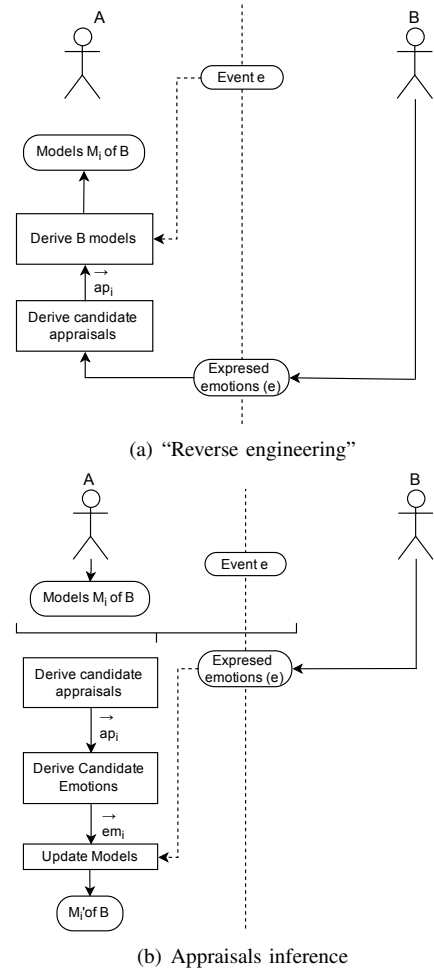


Fig. 1. Possible processes for an agent A to use agent B ’s expression on belief revision.

is no ambiguity to solve, but if both observations make the individual to think on different emotions then, according to [15], the observer may use some sort of strategy to find out what the observed individual is actually feeling. On the other hand, in [5] the authors perform studies that demonstrate that, by considering only facial expressions it is not possible to infer specific emotions if circumstances or situation are not evaluated. They argue that the situation can even determine the specific emotion chosen. If situation and expression, when judged individually, are congruent in the nature and magnitude of pleasure and arousal associated to them, then individuals tend to choose specific emotions guided by the situation. In contrast, if situation and expression are incongruent in the nature and magnitude of pleasure and arousal, then the facial information takes precedence when judging specific emotions.

III. DISAMBIGUATION AT A COGNITIVE LEVEL

Our approach is grounded in ToM [16] and appraisal theory [17], [18]. Starting from an event, an agent’s prediction about another’s emotional reaction depends on appraisals, which in turn are a byproduct of the agent’s ToM capacity to reason about the relationship between a situation and the other’s

concerns. We use PsychSim: a multiagent-based simulation tool [19], in order to simulate human social behavior and also the structures used are based in PsychSim modeling of the world. PsychSim is grounded in ToM, allowing us to build entities that can have goals, beliefs, and also mental models about other entities.

A. PsychSim modeling of the world

In PsychSim agents can have beliefs about the goals and behavior of other agents through fully specified decision theoretic models. An agent can also have models about itself. These quantitative recursive models include information about other agents' beliefs and expectations that allow, for example, to assess the impact of the agent's behavior on the self or other agents' goals and beliefs.

Basically an agent in PsychSim maintains its own beliefs about the world, goals and polices to achieve the goals. Through a *state* structure the agent keeps track of the values of the features that represent its subjective view of its 'True' state. An agent's *actions* are the actions an agent can chose to perform to modify the state of the world. The *goals* of an agent determine what are the motivations for some behavior, which is represented through a reward function that allows the agent to evaluate cost and benefits given a state of the world. Goals have weights that establish the relative preferences of goals. The agent selects an action by using *policies*. In PsychSim, policies can take several forms, but here we assume only lookahead policies whereby the agent projects the effects of its actions as well as how other agents will re-act. Thus this projection simulates the behavior of the other agents and the dynamics of the world for each action. This projection is bounded for N steps in the future. This bounded projection is designed to reflect the bounded reasoning that people perform in a real world. Finally an agent has also *beliefs* which are a representation of what an agent thinks is the state of the world. Beliefs have the same representation of the objective real world. An agent's beliefs about other agents are its "mental models" of other agents and they include the other agent's goals, beliefs and polices.

B. General steps

In order to describe the process for updating an agent's mental models, we will use a simple, illustrative situation: i) an agent A performs an action a , ii) then an agent B reacts to a with an expression E that A perceives and finally iii) the agent A updates its beliefs by using the emotions corresponding to the expression E and its beliefs of B. In order to determine how the mental models that agent A has about B are updated three main steps were performed:

- 1) Inference of the appraisal derivations patterns.
- 2) Inference of emotion categories.
- 3) Updating of the mental models.

1) *Deriving appraisal patterns:* For A to update its beliefs, and specifically its mental models about B starting from what it already thinks about B, at first it is required to find out what would be B's appraisals according to A's current beliefs.

In PsychSim an agent can project the effects of an action it performs N steps in the future. Assuming that the agents' turns alternate, then each step of the projection in the inference process contains:

- The actions $\vec{A}_{ag_i} = \langle a_1, a_2, \dots, a_\alpha \rangle$ an agent ag_i can perform for a given state s^t of the world.
- For each action the possible mental models $\vec{M}_{ag_i} = \langle m_1, m_2, \dots, m_\mu \rangle$ of the agent ag_i that can be considered.
- Finally, for each possible action a_i and mental model m_i to use for agent ag_i it is possible to get also different possible states $\vec{S}^{t+1} = \langle s_1^{t+1}, s_2^{t+1}, \dots, s_\gamma^{t+1} \rangle$. That is: if the action a_i performed according to mental model m_i does not produce more than one result for each feature of s^t (in other words, there are no probabilistic effects), then there is only one candidate state s^{t+1} with probability 1, otherwise the result is a probability distribution.

Then the total number of candidate next states of the world given an action (equivalent to the number of branches for this action) is determined by the number μ of mental models about agent B and the number γ of candidate states according to the action a . γ is independent of μ and its value is always the same regardless of the model m so the number of branches nb can be calculated as $nb = \mu * \gamma$. Considering each combination action-model-state as a "branch of projection" $b_k(a, m, s)$ starting from a given state s^t , a PsychSim agent can calculate appraisals as a byproduct of its normal forward projection, as Si et al. [9] argues. Specifically, the agent will calculate the different utilities associated to each "branch" and hence, different possible appraisal patterns and emotional reactions for agent ag_i for each new candidate state of the world. Here we consider four key appraisal dimensions that are inspired by Smith and Lazarus's [20] and Roseman's [21] theories of cognitive appraisal: *motivational congruence*, *motivational relevance*, *control* and *probability*. The model of appraisal proposed in [10] also supports the appraisal derivation in our approach, since it studies the coupling between the model of appraisals and PsychSim as a decision-theoretic framework. It is also based on Smith and Lazarus's theories of appraisal.

a) *Motivational congruence and relevance:* *Motivational relevance* expresses the extent to which an event is related to individual goals and *motivational congruence* represents the extent to which this event facilitates or inhibits individual goals [20]. In order to determine the *motivational congruence* and *motivational relevance* of each $b_k(a, m, s)$ we start from the concepts of *current utility* and *previous utility* used in [10]. *Previous utility* is the expected utility for agent B before agent A performs the action a (which is the event under analysis), and *current utility* is the expected utility for agent B after the action a is performed by agent A, reflecting its effect. These values determine the ratio of utility change (representing *motivational relevance*) and the direction of the change (representing *motivational congruence*).

b) *Control:* Also called "control potential", sets to what extent an individual can take control or influence the motive-relevant aspects of an event [21]. It means: to what extent

something can or cannot be done (by the individual under analysis or others), about the situation produced by an event. Returning to our generic example, the computation of *control* uses the lookahead of the agent A after performing a and, for each possible action of agent B on the frontier of each branch $b_k(a, m, s)$ it was determined if the correspondent utility was higher than the *previous utility* (the same used to calculate motivational congruence and relevance). The magnitude of control is then given by the proportion of “higher utilities” on the branches with respect to the number of branches. Note that the actions under analysis can be performed by the agent B itself or by any other agent, which means that B could seek instrumental social support from those other agents [9].

c) Probability: According to [21] this appraisal dimension indicates the probability of the situation produced by an event. Given the action a performed by agent A of the generic example, it is the probability of each branch conditioned on this action a : $Pr(b_k|a)$. The branch is determined by the mental model selected for agent B and one of the resulting states. In PsychSim the mental models that agent A has about agent B are organized in a probability distribution, as well as the candidate states for an action a and model m . They determine the probability of the branch. The dependency between m and the choice of the action, a , is independent of the dependency between the effect of the action, s , once a has been chosen. Equation 1 shows this result.

$$\begin{aligned} Pr(b_k|a) &= Pr(m, s|a) \\ &= Pr(m|a) \cdot Pr(s|a) \end{aligned} \quad (1)$$

2) Deriving emotion categories: In section III-B1 it is described how an appraisal pattern is obtained for each branch of projection. The probability of each appraisal pattern is given by the probability of the corresponding branch (see Equation 1) and hence, this will be also the probability of the emotion corresponding to the branch. Table I shows the mapping from an appraisal pattern to an emotion category which is mainly based in the mappings of [21] and [22]. We assume as ‘high’ those values above the medium value in the range of possible values and ‘low’ otherwise. The probability of each emotion category Ec_j is the sum of the probabilities of the corresponding appraisal patterns, which is equivalent to the probability of the corresponding branch, b_k (see Equation 2). Note that $Pr(Ec_j|b_k)$ will be 1 for any branch from where Ec_j was derived ($Ec_k = Ec_j$) and 0 when $Ec_k \neq Ec_j$. Note that multiple b_k are consistent with a given Ec_j , i.e. the same Ec_j can be derived from more than one b_k .

$$\begin{aligned} Pr(Ec_j) &= \sum_{k=1}^{nb} Pr(b_k, Ec_j) \\ &= \sum_{k=1}^{nb} Pr(Ec_j|b_k) \cdot Pr(b_k) \end{aligned} \quad (2)$$

Equation 3 shows how to calculate the probability of Ec_j for a model m , which is the sum of the probabilities of the

TABLE I
MAPPING FROM AN APPRAISAL PATTERN TO AN EMOTION CATEGORY

Appraisal Pattern for branch				Emotion
Motiva tional relevance = high	Motivational congruence = high			<i>joy</i>
	Motivational congruence = low	Control = high		<i>anger</i>
		Control = low	Probability = high	<i>sadness</i>
	Probability = low		<i>fear</i>	
Motivational relevance = low				<i>neutral</i>

branches that include m . In equation 3 $Pr(Ec_j|m, b_k)$ is the probability of Ec_j for a branch and model, and $Pr(b_k|m)$ is either 0 or 1. If it is 1 then m does not provide any more information about Ec_j than b_k already does.

$$\begin{aligned} Pr(Ec_j|m) &= \sum_{k=1}^{nb} Pr(Ec_j, m|b_k) \\ &= \sum_{k=1}^{nb} Pr(Ec_j|m, b_k) \cdot Pr(b_k|m) \end{aligned} \quad (3)$$

3) Updating models: The underlying idea in order to judge what an individual is feeling from a point of view of an observer is informed by Wallbott’s work [15] and J. M. Carroll and J. A. Russell’s work [5]. In our approach the observer starts from two distributions of possible emotion categories. One distribution represents what emotions Ec_ℓ the perceived expression of agent B suggests and to what extent; we will call it Θ . The other distribution represents the emotions derived from the appraisals for each possible model of the observer. The update of the observer’s beliefs will depend then on the comparison of these two distributions. For example, if the perception and the inference are the same then there is a high confidence in the current state of the observer’s beliefs and no update is made. On the contrary, if perception and inference are incongruent then perceptions will have more relevance, and hence, they will impact the updating of the observer’s beliefs.

The way to compute the update of the distribution over agent A’s mental models’ probabilities in relation to agent B is shown in Equations 4, 5, and 6.

$$Pr(m|\Theta) = \text{normalize}(Pr(\Theta|m) \cdot Pr(m)) \quad \text{Bayes rule} \quad (4)$$

$$= \text{normalize}\left(\prod_{\ell} Pr(\Theta_\ell|m) \cdot Pr(m)\right) \quad (5)$$

$$= \text{normalize}\left(\prod_{\ell} \left(\sum_j Pr(\Theta_\ell|m, Ec_j) \cdot Pr(Ec_j|m)\right) \cdot Pr(m)\right) \quad (6)$$

Let m be one of the possible mental models agent A has about agent B. Θ represents perceived emotional categories (distinct from Ec_j which represents true emotional categories). $Pr(Ec_\ell|\Theta)$ is the probability of emotion category Ec_ℓ in the perception but we refer to this emotion category just as Θ_ℓ in Equations 5 and 6 for simplification.

In equation 4, $Pr(\Theta|m)$ represents the likelihood of the perception given the selected mental model m for agent B. This expression is derived using Bayes’ rule.

In Equation 5, the probability of the perception is expressed in relation to Θ individual emotion categories¹. In Equation 6, $\Pr(\Theta_\ell|m, Ec_j)$ can be intuitively described as: the likelihood of perceiving the emotion category l given all possible true emotion categories j of model m . Note that, $\Pr(\Theta_\ell|m, Ec_j) = \Pr(\Theta_\ell)$ when $\ell = j$ (same emotion categories) and 0 otherwise.

Thus far we have shown how to update an agent’s beliefs. One way to update the perceptions would be to calculate the probability of the perception given the updated belief (see Equation 7 for reference). We return to this issue in the discussion section.

$$\begin{aligned} \Pr(Ec_j|\Theta) &= \text{normalize} \left(\prod_{\ell} \Pr(\Theta_\ell|Ec_j) \cdot \Pr(Ec_j) \right) \\ &= \text{normalize} \left(\prod_{\ell} \Pr(\Theta_\ell|Ec_j) \sum_m \Pr(Ec_j|m) \cdot \Pr(m) \right) \end{aligned} \quad (7)$$

IV. BOSS WORKER EXAMPLE

In order to illustrate our approach we show how the simple example of section III-B is implemented in PsychSim. The example represents a situation where a *worker* is told that he is fired. His *boss* must then find out if what he currently thinks about the *worker* is accurate or not according to the emotional expression of the *worker*.

Table II shows the possible models for each agent. Each agent can store multiple models of itself, some models are the actual, noted as ‘True’ models that the agent uses in its reasoning, while other models may be specified that are the models that other agents can use to reason about that agent. In PsychSim an agent’s beliefs about another agent can be represented as a discrete distribution of possible (mental) models of the other agent. In the case of the *boss*, he believes that the *worker* ‘loveswork’ with a probability of 0.4 and that the *worker* ‘hateswork’ with a probability of 0.6. Additionally, the *boss* has beliefs about the *worker*’s beliefs about the *boss*. Specifically, the *boss* ‘hateswork’ model of the *worker*, for example, is linked to the *worker* beliefs that he (the *boss*) behaves as an ‘efficientBoss’ with probability 0.2 and that he behaves as a ‘notEfficientBoss’ with probability 0.8. On the other hand the *worker* doesn’t store beliefs about the *boss* so he will use the *boss*’s ‘True’ model as a default.

Figure 2 depicts the actions each class of agent can perform (parallelograms), the features that belong to them or to the world (ellipses) and the models of each agent (documents). Each possible mental model is represented with dotted lines and it is associated with its corresponding reward functions. For example, in the case of the *worker* agent, when behaving under the model ‘loveswork’, he will try to set the feature ‘jobActive’ to True and he will try to maximize its feature ‘evaluation’. In PsychSim it is also possible to specify the dynamics of the actions that agents can perform. For example when the *boss* ‘fires’ the *worker*, in response the *worker* can

TABLE II
AGENTS’ MENTAL MODELS DISTRIBUTION

Agent	Possible models	Mental models
boss	True	worker :loveswork (0.4) worker: hateswork (0.6)
	efficientBoss	-
	notEfficient	-
worker	True	-
	hateswork	boss: efficientBoss (0.2) boss: notEfficientBoss (0.8)
	loveswork	-

TABLE III
RESULTS OF THE INFERENCE PROCESS OF THE AGENT ‘BOSS’ IN RELATION TO THE AGENT ‘WORKER’.

PROJECTION			
	boss belief: worker hateswork		boss belief: worker loveswork
boss action	worker action		
doNothing	*searchCandidateJob		*searchCandidateJob
*fire	argue	boss: efficientBoss	*argue
		boss: notEfficientBoss	
	*search Candidate Job	boss: efficientBoss	*searchCandidateJob
		boss: notEfficientBoss	
reduceWage	*searchCandidateJob		*searchCandidateJob
EMOTIONS			
Derived emotions	joy (0.6), fear (0.1), anger(0.1), sadness (0.1), neutral (0.1)		joy (0.1), sadness (0.1), anger(0.1), fear (0.4), neutral (0.3)
UPDATED MODELS			
	perception: fear (0.6), anger (0.4)		
boss modified beliefs	worker: loveswork (0.73) hateswork (0.27)		

even ‘argue’ to improve his evaluation and get better chances to recover his job, or ‘search for candidates jobs’. Also when the *boss* ‘reduces the wage’ of the *worker*, the *worker* can also ‘search for candidates jobs’. Additionally either for ‘fire’ or for ‘reduce wage’ the *worker* has an emotional expression in response that has associated a distribution of emotions categories. Naturally, and according to the rewards functions of each model of the *worker*, if the *worker* ‘loveswork’ he will try to ‘argue’ instead of search candidate jobs because to recover his job has much more importance to him than switch to another work. Similarly, if he ‘hateswork’ the action ‘searchCandidateJobs’ will have more value to him than the action ‘argue’.

Table III shows roughly the results of the process of inference that the *boss* performs in relation to what he believes about the *worker*. In this inference process both possible models of the *worker* are considered: ‘loveswork’ and ‘hateswork’, and also, all the information associated to them as their distributions of probabilities, their reward functions and their own beliefs. The actions that report the best utility for the *worker*, after the action of the *boss*, are marked with an asterisk. So for example, if the *boss* fires the *worker*, and the *worker* hates his work, then the *worker* will prefer searching for candidate jobs.

¹We assume that all emotion categories are independent

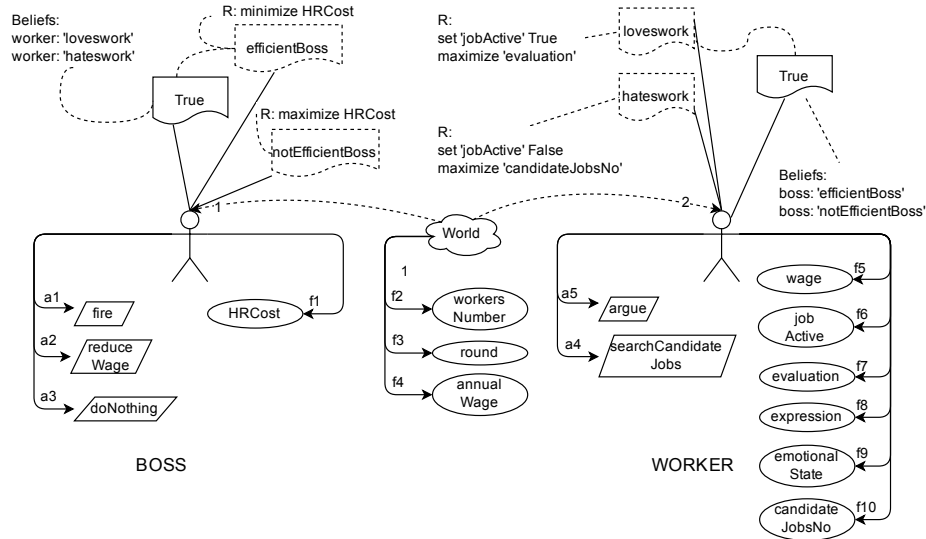


Fig. 2. Actions (parallelogram), features (ellipses) and models (document shapes) of entities in the boss-worker example.

Additionally, in the reasoning of the *boss*, when the *worker* “hateswork”, the *worker* would have different beliefs about the *boss*, that need to be considered independently as well as their probabilities. That is why for example, for the action “fire” of the *boss* and the action “argue” of the *worker*, one more step of the simulation is performed that generates two more branches. This inference process helps the *boss* to figure out what emotions the *worker* may experience after his action of firing him, being “fire” the *boss*’s best action to perform given the current state of the world. The *boss* finds that if the *worker* “hateswork” then the *worker* will feel joy with a probability of 0.6 and if the *worker* “loveswork” then the experienced emotion will be fear (0.4). We assign a small probability (0.1) to those emotions not inferred from the model since there is not a complete certainty of their absence. At the end of Table III the *boss* modified beliefs are shown. This update starts from a perception that gives to the emotion ‘fear’ and to emotion ‘anger’ probabilities 0.6 and 0.4 respectively. The last row of Table III shows the updates of the models. The model that fits better with the perception is “loveswork”, since one of the emotions inferred from it was also perceived.

V. DISCUSSION AND FUTURE WORK

The capability to model others can potentially benefit human computer interaction, as well as, more specifically, the design of agents that can interact with people and other emotional agents. Effective use of these models requires the ability to update them in light of ambiguous observations and conversely to disambiguate those observations. That has led us to explore the question of how predictions about another agent’s emotional reaction is integrated with perception of the expressions of emotion and how that integration refines those predictions and revises the observations.

In this work, we chose to leverage a Bayesian approach to the problem, adding functionality to an existing multi-

agent system to realize the approach. We focused mainly on the part of the problem whereby prediction and observation revises beliefs and only briefly touched on the question of how predictions might disambiguate observations. We proposed and implemented a three step method that follows a top-down mechanism to revise beliefs by integrating probabilistic inferences and probabilistic perception of the emotional expression.

Several questions remain. Our focus was largely on how ambiguous perceptions might update uncertain beliefs and we only briefly touched upon how strong prior beliefs might alter perception. One way to approach this, as we showed earlier is to assume the updated beliefs in turn influence perception, thereby not fully addressing what takes precedence: the judgment of the perception individually, the inference from the situation, or both? For example, one might instead model it as more of a simultaneous, mutual influence. Further the calculations assumed uncertainty is resolved at the level of the emotion category, but one might argue instead that the uncertainty is resolved at the level of appraisals or facial features which can lead to different underlying correspondences between predictions and perceptions. The next steps of research should also address how the system can be extended to a wider and more diverse set of emotions; how to extend the approach with models to derive emotions from the expression; and how to use this structures to update either the models of expression of other agents or the nature of the emotions actually expressed by other agents.

ACKNOWLEDGMENTS

The authors would like to thank the “Human and Social Behavior Models for Agent-Based Computational Economics” project: HUMBACE. GVA. PROMETEOII/2013/019. Thanks also to the Research and Development Support Programme of the Universitat Politècnica de València PAID-06-11 Project.

REFERENCES

- [1] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique." *Psychological Bulletin*, vol. 115, pp. 268–287, 1994.
- [2] S. S. Tomkins, *Affect, Imagery, Consciousness: The Positive Affects*. Springer, 1962, vol. 1.
- [3] —, *Affect, Imagery, Consciousness: The Negative Affects*. Springer, 1963, vol. 2.
- [4] C. E. Izard, "Innate and universal facial expressions: evidence from developmental and cross-cultural research." *Psychological Bulletin*, vol. 115, pp. 288–299, 1994.
- [5] J. M. Carroll and J. A. Russell, "Do facial expressions signal specific emotions? Judging emotion from the face in context." *Journal of personality and social psychology*, vol. 70, no. 2, p. 205, 1996.
- [6] S. Hareli and U. Hess, "What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception," *Cognition and Emotion*, vol. 24, no. 1, pp. 128–140, 2010.
- [7] C. Melo, P. Carnevale, and J. Gratch, "The Effect of Virtual Agents' Emotion Displays and Appraisals on People's Decision Making in Negotiation," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, Y. Nakano, M. Neff, A. Paiva, and M. Walker, Eds. Springer Berlin Heidelberg, 2012, vol. 7502, pp. 53–66.
- [8] A. Bubic, D. Y. Von Cramon, and R. I. Schubotz, "Prediction, cognition and the brain," *Frontiers in Human Neuroscience*, vol. 4, no. 25, 2010.
- [9] M. Si, S. C. Marsella, and D. V. Pynadath, "Modeling appraisal in theory of mind reasoning," *Autonomous Agents and Multi-Agent Systems*, vol. 20, no. 1, pp. 14–31, 2010.
- [10] D. V. Pynadath, M. Si, and S. C. Marsella, *Modeling theory of mind and cognitive appraisal with decision-theoretic agents*. Oxford University Press, 2011, pp. 70–87.
- [11] Y. Trope, "Identification and inferential processes in dispositional attribution," *Psychological review*, vol. 93, no. 3, pp. 239–257, 1986.
- [12] J. S. Uleman, "Consciousness and Control The Case of Spontaneous Trait Inferences," *Personality and Social Psychology Bulletin*, vol. 13, no. 3, pp. 337–354, 1987.
- [13] G. A. Van Kleef, C. K. De Dreu, and A. S. Manstead, "The interpersonal effects of anger and happiness in negotiations." *Journal of personality and social psychology*, vol. 86, no. 1, p. 57, 2004.
- [14] —, "Supplication and appeasement in conflict and negotiation: The interpersonal effects of disappointment, worry, guilt, and regret." *Journal of personality and social psychology*, vol. 91, no. 1, p. 124, 2006.
- [15] H. G. Wallbott, "Faces in context: The relative importance of facial expression and context information in determining emotion attributions." in *Facets of Emotion*, K. R. Scherer, Ed. Hillsdale: Lawrence Erlbaum, 1988.
- [16] A. Whiten, *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. Basil Blackwell Oxford, 1991.
- [17] K. R. Scherer, "Toward a dynamic theory of emotion: The component process model of affective states," *Geneva Studies in Emotion and Communication*, vol. 1, no. 1, pp. 1–98, 1987.
- [18] N. H. Frijda, *The Emotions*, ser. Studies in Emotion and Social Interaction. Cambridge University Press, 1987.
- [19] S. C. Marsella, D. V. Pynadath, and S. J. Read, "PsychSim: Agent-based modeling of social interactions and influence," in *Proceedings of the International Conference on Cognitive Modeling*, vol. 36, 2004, pp. 243–248.
- [20] C. A. Smith and R. S. Lazarus, "Emotion and Adaptation," in *Handbook of Personality: Theory and Research*, L. A. Pervin, Ed., 1990, pp. 609–637.
- [21] I. J. Roseman, *A model of appraisal in the emotion system: Integrating theory, research, and applications*. Oxford University Press, pp. 68–91.
- [22] S. C. Marsella and J. Gratch, "EMA: A process model of appraisal dynamics," *Cognitive Systems Research*, vol. 10, no. 1, pp. 70–90, 2009.