Emotion-mapped Robotic Facial Expressions based on Philosophical Theories of Vagueness

Phil Serchuk[1,2] Ehud Sharlin[2] Martin Lukac[3] Marek Perkowski[3]
[1] Department of Philosophy, University of Calgary, pserchuk@ucalgary.ca
[2] Department of Computer Science, University of Calgary, ehud@cpsc.ucalgary.ca
[3] Department of Electrical and Computer Engineering, Portland State University, mperkows@ee.pdx.edu, lukacm@cecs.pdx.edu

ABSTRACT
As the field of robotics matures robots will need some method of displaying and modeling emotions. One way of doing this is to use a human-like face on which the robot can make facial expressions corresponding to its emotional state. Yet the connection between a robot's emotional state and its physical facial expression is not an obvious one: while a smile can gradually increase or decrease in size, there is no principled method of using boolean logic to map changes in facial expressions to changes in emotional states. We give a philosophical analysis of the problem and show that it is rooted in the vagueness of robot emotions. We then outline several methods that have been used in the philosophical literature to model vagueness and propose an experiment that uses our humanoid robot head to determine which philosophical theory is best suited to the task.

INTRODUCTION
It has been argued that the ability to display emotions on a human-like face is both an important and necessary step in making robots and computer agents more accessible to the general public [1, 7]. The emotional model for a robotic face will have two key components: a mapping of emotional states to facial positions and a method of transitioning between different pairs of emotional states and facial positions. For example, we map 'happy' on our robot to a widening of its face and eyes and a slight opening of its mouth: in other words, a smile. We say the robot is 'sad' when its eyes fall and face narrows. Some method is now needed of transitioning between 'happy' and 'sad' that both displays the transition on the robot's face and captures the emotional state of the robot throughout the transition.

We begin by introducing our notation and the philosophical problem of vagueness. We then show how, in boolean logic, there is no principled method of selecting a particular facial position to use as the boundary between two different emotional states. The problem arises because robot emotions, along with many linguistic predicates in natural language, are vague. Philosophers have studied paradoxes of vagueness since antiquity and while they are no closer to solving problems of vagueness than computer scientists, an examination of the philosophical problem of vagueness will allow researchers to better model emotions and their corresponding facial expressions in a robot face.

PRELIMINARIES AND NOTATIONS
The robot head we are using consists of many parts, each of which is controlled by a servo. Let $\hat{v}$ be a vector that contains the variables which denote the servo's position. We use $P_{\hat{v}}$ to denote the head's position when the servos are set to the positions contained in $\hat{v}$. For example, $\hat{v} = \{100, 90, 40\}$ would set $face_width = 100$, $eyelid_width = 90$, and $mouth_open = 40$. Although our particular robot head contains up to 16 servos, we have combined some and omitted others from $\hat{v}$ to simplify our examples. We define the following vectors:

$\hat{v1} = \{100, 100, 40\}$
$\hat{v2} = \{85, 100, 40\}$
$\hat{v3} = \{50, 100, 40\}$
$\hat{v4} = \{15, 100, 40\}$
$\hat{v5} = \{0, 100, 40\}$

Each emotional state is denoted by a boolean variable S_α, whose value is returned by the function t. We use S_h to denote the emotional state 'happy'. The robot is 'happy' if and only if $t(S_h) = 1$; the robot is 'not happy' if and only if $t(S_h) = 0$. We use $\neg$ to denote logical negation; thus, $\neg S_h$ denotes the state 'not happy'. Because we are initially working in boolean logic, we assume that if the robot is 'not happy' it is 'sad'; thus, the robot is 'sad' when $t(S_h) = 0$, or $t(\neg S_h) = 1$. To denote the robot's emotional state given its facial position, expressed as a vector $\hat{v}$, we use $S_{\alpha_{\hat{v}}}$.

Events cause, among other things, the robot's facial position to move. For example, as negative events occur the robot's smile will slowly disappear. As is the case with humans, multiple facial positions can correspond to the same emotion. For example, both $\hat{v1}$ and $\hat{v2}$ map to state S_h, meaning $t(S_{h_{\hat{v1}}}) = 1$ and $t(S_{h_{\hat{v2}}}) = 1$. We let $\hat{v4}$ and $\hat{v5}$ map to state 'sad', $\neg S_h$, which means

$t(\neg S_{h_{\hat{v4}}}) = 1$ and $t(\neg S_{h_{\hat{v5}}}) = 1$.

THE VAGUENESS OF ROBOT EMOTIONS

Suppose we want to build a robot office assistant that, among other things, delivers mail to office employees. We define its facial position as $P_{\hat{v6}}$, where $\hat{v6} = \{i, 100, 40\}$. Note that when $i = 100$, $\hat{v6} = \hat{v1}$, and $t(S_{h_{\hat{v6}}}) = 1$. Each time the robot successfully completes a task we increase i to make the robot look happier, and each time the robot fails to complete a task we decrease i to make the robot look sadder. For example, if the robot spills coffee on a human we might set $i = i-70$, but if the robot delivers an envelope to the wrong employee we might only set $i = i - 1$.

The Boundary Between S_h and $\neg S_h$

The fundamental problem with this model is that we cannot determine the i that marks the location where the robot changes from state S_h to state $\neg S_h$. Suppose the face is in position $P_{\hat{v3}}$. We are just as justified in holding that position $P_{\hat{v3}}$ corresponds to state $\neg S_h$ as we are in holding that position $P_{\hat{v3}}$ corresponds to state S_h. We do not want to arbitrarily assign $P_{\hat{v3}}$ to S_h or $\neg S_h$, as to hold that position $P_{\hat{v3}}$ is the precise cut-off between emotional states S_h and $\neg S_h$ is to hold that while some large number of negative events may occur to move the robot's face from position $P_{\hat{v1}}$ to $P_{\hat{v3}}$, the robot's emotional state remains constant at S_h; yet if one small, seemingly insignificant event causes the robot's face to move past position $P_{\hat{v3}}$ and toward $P_{\hat{v4}}$, the robot's emotional state will suddenly change from S_h to $\neg S_h$. This is clearly counterintuitive as, among healthy humans, small, insignificant actions should not cause sudden changes in emotional states.

Suppose we use $P_{\hat{v3}}$ as the boundary between S_h and $\neg S_h$. Recall that the face of the office robot is in position $P_{\hat{v6}}$ and suppose that, at a given point during the day, $i = 85$: since $\hat{v6} = \hat{v2}$ we know the robot is 'happy'. If the robot spills coffee on an employee then $i = 85 - 70 = 15$ and $\hat{v6} = \hat{v4}$; the robot becomes 'sad'. This result conforms to our intuition that if a person is happy and they make a big mistake, such as spilling coffee on a co-worker, they will become unhappy.

Now suppose $i = 85$ and the robot brings a piece of mail to the wrong employee: $i = 85 - 1 = 84$. This result also conforms to our intuitions: if a person is happy and they make a small mistake, such as bringing a piece of mail to the wrong person, they don't stop being happy. Now suppose this occurs 36 times: $i = 84 - 1 = \cdots = 50 - 1 = 49$. When the robot incorrectly delivers the mail for the 35^{th} time it is still happy, because $i = 50$, although its smile has nearly disappeared; however, once the robot delivers mail to the wrong person for the 36^{th} time, its mood immediately changes from 'happy' to 'sad', even though its physical appearance — the diminished smile that is visible when $i = 50$ — has changed by an imperceptibly small amount, namely by one unit, such that $i = 50-1 = 49$. The problem is that while the physical transition between the appearance of 'happy' and the appearance of 'sad' is gradual, as it is with humans, the transition between S_h and $\neg S_h$ is not: despite its physical appearance, the robot will believe it is as happy when $i = 51$ as it is when $i = 85$. When i is decremented by one unit the change in facial position will be hardly noticable, yet its mood instantaneously changes from 'happy' to 'sad': from S_h to $\neg S_h$.

The Logic

The problem in the above example lies in our belief that one insignificant event does not make the difference between emotional states. The linguistic vagueness that philosophers typically study works similarly: words like 'bald' are said to be vague because one hair does not seem to make the difference between baldness and non-baldness. Let $S_{h_{\hat{v6}_i}}$ indicate that the robot is in state S_h and in position $P_{\hat{v6}}$, given the value of i in $\hat{v6}$. Thus, $t(S_{h_{\hat{v6}_{100}}}) = 1$, since when $i = 100$, $\hat{v6} = \hat{v1}$. In boolean logic, we symbolize our belief that one insignificant event does not make the difference between S_h and $\neg S_h$ as follows: $\forall n(S_{h_{\hat{v6}_n}} \supset S_{h_{\hat{v6}_{n-1}}})$. Yet given this, we can prove that there is no i that marks the cut-off between S_h and $\neg S_h$ (see Figure 1).

1	$S_{h_{\hat{v6}_{100}}}$	
2	$\forall n(S_{h_{\hat{v6}_n}} \supset S_{h_{\hat{v6}_{n-1}}})$	
3	$S_{h_{\hat{v6}_{100}}} \supset S_{h_{\hat{v6}_{99}}}$	$\forall$E, 2
4	$S_{h_{\hat{v6}_{99}}}$	$\Rightarrow$E, 1, 3
$\vdots$	$\vdots$	
2,000	$S_{h_{\hat{v6}_1}}$	$\Rightarrow$E, $\vdots$
2,001	$S_{h_{\hat{v6}_1}} \supset S_{h_{\hat{v6}_0}}$	$\forall$E, 2
2,002	$S_{h_{\hat{v6}_0}}$	$\Rightarrow$E, (2,000), (2,001)

Figure 1. Proof: if the robot is 'happy' when in $P_{\hat{v6}_{100}}$, then it is 'happy' when in $P_{\hat{v6}_0}$

Thus, having started with an obviously true premise, namely $S_{h_{\hat{v6}_{100}}}$, we can conclude that when $i = 0$ the robot is still 'happy'; $t(S_{h_{\hat{v6}_0}}) = 1$. We could construct a similar argument, starting with $\neg S_{h_{\hat{v6}_0}}$ and assuming $\forall n(\neg S_{h_{\hat{v6}_n}} \supset \neg S_{h_{\hat{v6}_{n+1}}})$, to show that when $i = 100$ the robot is still 'not happy': $t(\neg S_{h_{\hat{v6}_{100}}}) = 1$. Because we are in boolean logic, where a robot that is 'not happy' is 'sad', we have been able to prove two seemingly contradictory facts. We proved both that the robot is 'happy' when $i = 0$ and its facial expression is 'sad', and that the robot is 'sad' when $i = 100$ and its facial expression is 'happy'. While boolean logic requires that we find some vector, and thus some i, to use as the boundary between S_h and $\neg S_h$, philosophers have developed sev-

eral theories of vagueness that we propose will resolve this contradiction and properly model robot emotions.

PHILOSOPHICAL THEORIES

Typically, philosophical theories of vagueness give both a metaphysical account of the phenomenon of vagueness and an account of our linguistic use of vague predicates. While our work is focused on modeling emotions in practical applications, the metaphysical content of these theories should not be disregard: as the field of human-robot interaction matures it is likely that scholarship will focus not only on the results achieved by a given method but on the correctness of using that method in the first place. In this section we present an overview of several theories of vagueness that seem particularly well-suited for the task of modeling robot emotions and controlling facial expressions; it is not our intent to give a detailed account of the philosophical literature. References for further reading are provided.

3-Valued Logics

For any variable P in boolean logic, either P is true or P is false. By introducing additional truth values, many-valued logics allow P to take other values. Many-valued logics are extensions of classical logic and always have 'true' and 'false' as truth values which behave, in relation to one another, as they would in classical logic [3, p. 5]. We use a Łukasiewicz 3-valued logic, which has the following truth values: $[0, \frac{1}{2}, 1]$. A truth value of $\frac{1}{2}$ represents an indeterminable truth value that is assigned to that which is possible and exists between 'the true' and 'the false'; that is, $\frac{1}{2}$ is truer than what is false but falser than what is true [8]. Given two variables P and Q, negation, conjunction, disjunction, and implication are defined thusly:

$t(\neg P) = 1 - t(P)$
$t(P \wedge Q) = min(t(P), t(Q))$
$t(P \vee Q) = max(t(P), t(Q))$
$t(P \rightarrow Q) = min(1, 1 - t(P) + t(Q))$

To model emotions in a 3-valued logic, we first determine which positions *clearly* correspond to S_h and which *clearly* correspond to $\neg S_h$. There are two ways that a 3-valued logic models states that are neither *clearly* S_h nor *clearly* $\neg S_h$: on the *truth gap* theory, these borderline cases are neither S_h nor $\neg S_h$ while on the *truth glut* theory they are both S_h and $\neg S_h$. On both *truth gap* and *truth glut* theories we have the following truth assignments: $t(S_{h_{\hat{v6}_{100}}}) = 1$, $t(S_{h_{\hat{v6}_{50}}}) = \frac{1}{2}$, and $t(S_{h_{\hat{v6}_0}}) = 0$. Recall that in section 1 the state 'sad' was said to be equivalent to the state $\neg S_h$: in boolean logic, if the robot is not in state S_h, such that $t(S_h) = 1$, then the robot must be in state $\neg S_h$, such that $t(S_h) = 0$. On the 3-valued approach, however, a robot that is not in state S_h need not be in state $\neg S_h$: when $t(S_h) = \frac{1}{2}$ the robot is not in state S_h nor is it in state $\neg S_h$. For more information on 3-valued logics and vagueness, see [8],[9], and [10].

Epistemicism

Epistemicism [11] holds that words like 'tall' and 'bald' have sharp boundaries that are necessarily unknowable to us. If we extend this view to emotions like 'happy', we must hold that when the robot is in $P_{\hat{v3}}$ it is really either in state S_h or $\neg S_h$ — but we can never know which state it truly belongs in. Of course, we can program the robot to treat $P_{\hat{v3}}$ as though corresponds to S_h, but this will yield the same sudden change in emotional state that we are trying to avoid. The strengths of this theory are primarily metaphysical, and the only practical advantage it offers is that it allows us to model vagueness, and emotions, in boolean logic.

To model emotions according to epistemic principles, we need to hold that there are precise facial expressions — some of which are unknown to us — that correspond to the robot appearing happy; all other facial expressions correspond to the robot looking unhappy. Thus, $P_{\hat{v1}}$ corresponds to the robot looking happy and $P_{\hat{v5}}$ corresponds to the robot looking unhappy. Yet according to the epistemicist, we have no way of knowing what the actual position of the cut-off is: we only know that it exists. Because epistemicism was developed for linguistic vagueness, many complications arise when attempting to model robot emotions using this theory. One possibility is to have the robot inform the user that it does not know whether it is happy or sad when in $P_{\hat{v3}}$; another is to have the robot stop expressing emotion altogether when it is in indeterminate positions like $P_{\hat{v3}}$ and continue expressing emotions when it returns to a position where its emotional state is clear, such as $P_{\hat{v2}}$.

Fuzzy Logic

Fuzzy logic has been proposed by philosophers and used by computer scientists to model linguistic vagueness; computer scientists have already developed several emotional models based on fuzzy logic [4, 5]. Fuzzy logic is an infinitely-valued logic, with truth values represented on the interval of real numbers $[0, 1]$. Variables, and in this case emotional states, are represented by fuzzy sets and objects in the domain are members of each set to varying degrees; the degree to which a particular variable belongs in the set **TRUE** is a variable's degree of truth. Negation, conjunction, and disjunction are defined as they are in 3-valued logic, and implication typically is as well, although other definitions are sometimes used. Thus, we initially know that $t(S_{h_{\hat{v6}_{100}}}) = 1$ and $t(S_{h_{\hat{v6}_0}}) = 0$. We map the other values of i to truth values using a membership function. Suppose the robot is *clearly* happy when $i > 80$ and *clearly* unhappy when $i < 30$. One possible membership function is:

if $i < 30$ then $t(S_{h_{\hat{v6}_i}}) = 0$
if $30 \leq i \leq 80$ then $t(S_{h_{\hat{v6}_i}}) = \frac{i-30}{50}$
if $i > 80$ then $t(S_{h_{\hat{v6}_i}}) = 1$

Using this function, the robot's emotional state changes along with its smile. When $i = 70$ the robot is 'happy'

to degree 0.8 and 'not happy' to degree 0.2, $t(S_{h_{\hat{v6}_{70}}}) = 0.8$; when $i = 60$ the robot is 'happy' to degree 0.4 and 'not happy' to degree 0.6, $t(S_{h_{\hat{v6}_{60}}}) = 0.4$. More information on fuzzy logic and fuzzy set theory can be found in [2] and [6].

USER STUDY: SELECTING A THEORY

Because we are not presently interested in the metaphysical claims of the theories described in the previous section, we must use some other criterion to determine which theory is best suited for mapping facial expressions to a robot's emotional state. Each theory of vagueness treats positions like $P_{\hat{v3}}$ differently: on the 3-valued approach $P_{\hat{v3}}$ corresponds either to both S_h and $\neg S_h$ or to neither S_h nor $\neg S_h$, while on the fuzzy logic approach $P_{\hat{v3}}$ partially corresponds to S_h and partially corresponds to $\neg S_h$. On the epistemic view $P_{\hat{v3}}$ corresponds to either S_h or $\neg S_h$, but it is impossible for us to know which. We propose conducting experiments with a robot face to test user's perceptions of emotions that are not clearly 'happy' nor clearly 'sad' to see how they perceive this borderline area.

Our experiment will begin with the head engaging in a scripted one-way interaction with the user. The user will be filling out a form in an office setting, with the robot seated across the table; note that only the upper body is visible to the user. At some point an event will occur that causes the robot to enter state S_h and move to position $P_{\hat{v1}}$; recall that $P_{\hat{v1}}$ makes the robot's face look as 'happy' as possible. The user will be explicitly informed that, when in this position, the robot is happiest. An event will then occur that causes the robot's face to move to position $P_{\hat{v5}}$ and enter state $\neg S_h$; recall that $P_{\hat{v5}}$ makes the robot's face look as 'unhappy' as possible. Again, the user will be explicitly informed that, when in this position, the robot is least happy.

We then let the user fill out the forms as a distraction task. Once a brief period of time has passed, the phone will ring and the robot head will answer, using a headset. While the user will only be able to hear the voice of the robot speaking, they will hear a muffled voice to indicate that someone is speaking to the robot over the phone. At various points in the phone call the robot's face will move to positions between $P_{\hat{v1}}$ and $P_{\hat{v5}}$; recall that these positions are between positions that make the robot appear *clearly* happy and *clearly* unhappy. Each time the face takes a new position the scenario will 'pause' and a facilitator will ask the user to describe the emotional state of the robot. The user will be able to choose one of the following options: 'happy', 'not happy', 'happy and not happy', 'neither happy nor not happy', 'either happy or not happy, but unsure of which', or 'partially happy and partially not happy'. The first two options conform to a traditional emotional model, while the last four are used to represent the theories of vagueness that were described in the previous section. We hypothesize that users will choose one of the first two options when the robot's state is obvious; when the state is difficult to determine, we believe users will choose one of the last four options. If a significant percentage of users choose one of the last four options when the robot is in an intermediate state, then we will have evidence indicating which theory of vagueness can be best used to control robot emotions and facial expressions.

We hope to conduct this test using at least the emotions 'happy' and 'surprised'. The scenario for 'surprised' will differ from the previously described scenario for 'happy'. If the data shows that users associate intermediate positions with a given theory of vagueness, then we will use literature on that theory to develop a more detailed model of robot emotions.

CONCLUSION

In this paper we reported our efforts to model synthetic emotions in a humanoid robot head using facial expressions based on philosophical theories of vagueness. We begun by highlighting the importance of accurate and valid emotional states in robotic interfaces and how these might be expressed through facial expressions. We argued that vagueness is a practical challenge when attempting to model convincing and interactive robotic emotions based on facial expressions. We presented three philosophical theories of vagueness that might be used in accurately modeling robotic emotions and facial expressions, and outlined a user study that will enable us to evaluate and compare the usability and effectiveness of these theories. We are currently applying the methodologies presented in the paper using a humanoid robot head that serves as our testbed for implementing the different facial expression models and for conducting the proposed user study.

ACKNOWLEDGMENTS

We'd like to thank Dr. Svetlana Yanuskevich from the University of Calgary Biometrics Lab, Andrew Read from the University of Calgary Science Workshop, and Brian Scowcroft from the University of Calgary Department of Computer Science.

REFERENCES

1. C. L. Breazeal. *Designing Sociable Robots*. MIT Press, 2002.
2. J. J. Buckley and E. Eslami. *An Introduction to Fuzzy Logic and Fuzzy Sets*. Physica-Verlag, 2002.
3. S. Gottwald. *A Treatise on Many-Valued Logics*. Studies in Logic and Computation. Research Studies Press Ltd., 2001.
4. L. C. Jain, H.-N. Teodorescu, and A. Kandel, editors. *Soft Computing in Human-Related Sciences*. CRC Press, 1999.
5. R. Langari and J. Yen. *Fuzzy Logic: Intelligence, Control, and Information*. Prentice-Hall Inc., 1999.
6. H. T. Nguyen and E. A. Walker. *A First Course in Fuzzy Logic*. CRC Press, 1997.
7. D. A. Norman. *Emotional Design*. Basic Books, 2004.
8. N. Rescher. *Many-valued Logic*. McGraw-Hill, 1969.
9. M. Tye. Vague objects. *Mind*, 99(396):535–57, 1990.
10. M. Tye. Vagueness: Welcome to the quicksand. In T. Horgan, editor, *Spindel Conference: Vagueness*, pages 1–22. University of Memphis, 1995.
11. T. Williamson. *Vagueness*. Routledge, 1994.