Applying Probabilistic Programming to Affective Computing

Desmond C. Ong, Member, IEEE Computer Society,
Harold Soh, Member, IEEE Computer Society,
Jamil Zaki, and Noah D. Goodman

Abstract—Affective Computing is a rapidly growing field spurred by advancements in artificial intelligence, but often, held back by the inability to translate psychological theories of emotion into tractable computational models. To address this, we propose a probabilistic programming approach to affective computing, which models psychological-grounded theories as generative models of emotion, and implements them as stochastic, executable computer programs. We first review probabilistic approaches that integrate reasoning about emotions with reasoning about other latent mental states (e.g., beliefs, desires) in context. Recently-developed probabilistic programming languages offer several key desiderata over previous approaches, such as: (i) flexibility in representing emotions and emotional processes; (ii) modularity and compositionality; (iii) integration with deep learning libraries that facilitate efficient inference and learning from large, naturalistic data; and (iv) ease of adoption. Furthermore, using a probabilistic programming framework allows a standardized platform for theory-building and experimentation: Competing theories (e.g., of appraisal or other emotional processes) can be easily compared via modular substitution of code followed by model comparison. To jumpstart adoption, we illustrate our points with executable code that researchers can easily modify for their own models. We end with a discussion of applications and future directions of the probabilistic programming approach.

Index Terms—Affective Computing, Artificial Intelligence, Emotion Theory, Modeling Human Emotion

1 INTRODUCTION

Achieving a human-like understanding of emotions is a holy grail of affective computing. The ideal affective computer has to correctly identify a user’s emotion based on behavioral cues and contextual information, reason about how its actions may affect the user’s emotions, and choose its responses accordingly. Affective agents deployed in a variety of applications, such as in tutoring [7], social robotics [77], or other human-computer interactions [53], aim to achieve this frontier of emotional understanding. Research in recent years has made important strides towards this goal, producing hundreds of papers on applying machine learning techniques to recognizing emotions (see [5], [30], [33], [90], for reviews). The success of these efforts is due in large part to recent developments in deep learning algorithms and computational power [38], coupled with the availability of larger datasets [14]. However, though these machine learning models may demonstrate excellent performance at emotion recognition, they nevertheless fall short of being able to do true reasoning [59]. They are unable to carry out counterfactual and hypothetical reasoning, provide causal attributions and explanations, or incorporate contextual knowledge into their inferences. And they usually do not generalize well outside the dataset they are trained on. By merely performing pattern recognition, these machine learning models achieve perception, but fall short of cognition about affect.

In contrast to the data-driven, machine learning approach, an alternative approach with a much older history in affective computing favors constructing theory-based models, such as emotion architectures (see [44], [50] for reviews). These models are inspired by psychological theories of emotion [18], [72], and focus in detail on modelling the components and the computations that go into emotional processes. For example, many such models focus on computationally defining the cognitive evaluations of experienced situations—appraisals—that give rise to emotions [4], [49], [60]. Other models also focus on modelling how emotions influence cognition and subsequent behavior [3], [47]. However, these models tend to be hand-tuned to specific theories and specific contexts—for example, the model in [60] contains many rule-based appraisals that give rise to emotions, but is not able to learn new appraisals or modify existing appraisal rules for new contexts—and thus they are unable to scale well to the complexity of larger, naturalistic datasets. Another limitation is that these models usually do not specify how one goes from naturalistic data, such as a pixel-level rep-
representation of a smiling face, to a representation of emotion: The exact "transformation" from an emotion to a visually-observable muscular configuration may be scientifically uninteresting (except perhaps for the fact that such a reliable and valid mapping does or does not exist), yet is still important to an engineer wishing to build applications that are sensitive to affect.

These theory-driven models differ not only in their theoretical assumptions and content, but also in the details of their implementation, such that it is problematic to compare theory side-by-side [44], [50]. To add to this difficulty, these theory-driven approaches tend to have large barriers to adoption, as most of them use specific architectures or systems that may be difficult for other researchers to adopt. For example, the EMA model [49] and Marinier and colleagues’ [47] model are built on top of SOAR [54], and a researcher who wishes to contribute to these lines of work will have to learn an unfamiliar set of syntax and conventions—although we note that some researchers have advocated for settling on a unified architecture [63]. Other research groups may implement their models in custom-built systems [3], [13], which may not be easily available for other researchers to build upon.

By contrast, deep learning and other machine learning approaches have experienced a meteoric rise in popularity and adoption over the past few years. This is due in large part to readily-available, open-source, deep learning tools written in beginner-friendly programming languages like Python and R, along with accessible learning materials (tutorials, online courses) and a large and active community (on programming forums and blogs like Stack Overflow and Medium respectively). An ideal approach is one that "democratizes" affective computing by being easier to adopt, increasing the speed of research and infrastructural developments, while still maintaining the scientific rigor and experimentation of theory-based approaches.

In this paper, we propose a probabilistic programming approach to affective computing that marries the strengths of theory-driven approaches with that of data-driven approach. Probabilistic programming is a modeling paradigm by which one can specify theories of emotion using probabilistic, generative models [22], [24]. We can explicitly represent uncertainty in emotion theory, randomness in emotional phenomena, or even incomplete knowledge about an agent’s mental state, as programs that contain some degree of randomness. Because probabilistic programs are modular and can be composed to form more complex programs, we can focus on modeling at different levels of abstraction [50], [63]. Having a hierarchy of abstract representations also allows probabilistic programs to learn context-specific knowledge (i.e., to the specific example) as well as knowledge that can generalize to other scenarios [74]. One can use a probabilistic programming framework to test different emotion theories (e.g., of appraisal) by substituting modular chunks of code and testing which theories best fit experimental data. Modularity also allows integration of emotion with other high-level theories of psychological phenomena, such as mental states and motivation [59], [68].

In probabilistic programming, model specification is orthogonal to learning and inference in these models. This separation allows the affective computing modeler to focus on specifying the model rather than on inference and optimization methods, much like how so-called "high-level" programming languages abstract away the workings of machine code from the programmer. Such abstraction lowers the barrier to adoption by making languages easier to learn, while simultaneously making them more efficient. Because inference in probabilistic programming is orthogonal to model specification, infrastructural development of inference algorithms can proceed in parallel [64]. Indeed, many modern probabilistic programming languages leverage existing deep learning, optimization, and inference libraries, which allows efficient and scalable learning from large datasets [76]. The modeller can simply specify the model, give it data, and press "run", so to speak, relying on general-purpose inference implemented within the probabilistic programming language. Our claim is that probabilistic programming combines the strengths of the data-driven and theory-driven approaches in affective computing described above: It allows researchers to build psychologically-grounded models—along with hypothesis testing and scientific experimentation—with the infrastructure to learn efficiently from and do inference over large data. And it provides a common platform with a low barrier to entry that will encourage integration among existing approaches in the field.

We begin the paper by introducing the intuitive theory approach to understanding reasoning about emotion, and how this approach has been formalized in the computational cognitive science literature using probabilistic methods—specifically, probabilistic graphical models. Next, we introduce probabilistic programming by discussing two recent high-impact examples of applying probabilistic programming to model human cognition. With this background, we then discuss implementing a computational model of emotion using probabilistic programming, by providing worked example code using a re-analysis of a small multimodal dataset [58]. We illustrate how one can model components such as appraisal and emotion recognition from faces. We also demonstrate reasoning capabilities that this probabilistic programming approach has over previous approaches: For example, it can generate novel emotional faces given a new situation. Finally, we end by discussing the boundaries of this approach, as well as the long-term promise of this approach, such as to modeling emotion generation and understanding complex intentional emotional displays.

2 IMPLEMENTING INTUITIVE THEORIES AS PROBABILISTIC MODELS

People have an intuitive understanding of the world around them. The average person may not be able to write down Newton’s laws, but, upon witnessing a thrown baseball, is able to intuitively predict the ball’s trajectory and where it would land [19], [51]. This intui-

---

1For example, object-oriented programming was a major abstraction that allowed complex data structures (objects) with their own functions.
tive understanding also extends to making sense of other people. If we see a roommate walk out of their room, pause, and turn back, we intuitively start generating possible hypotheses about their behavior—perhaps they forgot something and went back to retrieve it—even though these inferences are made on sparse and incomplete information. In daily life, we effortlessly make such inferences about other people’s thoughts, feelings, and even personality and other traits [21], [31], [66]. Such “intuitive physics” and “intuitive psychology” are made possible by intuitive theories that consist of a structured ontology of concepts—gravity, air resistance, and velocity; goals, personality, and behavior—and the causal relationships between these concepts [20], [28]. This knowledge allows us to make predictions, complex inferences, and offer explanations about the observed world. If we know that Bob wants to eat (goal), we can predict that he will go look for food (behavior). Conversely, if we see that Bob has gone to the cupboard and returns with some potato chips, we can also infer that he knew that there were snacks in the cupboard (Bob’s beliefs about the world), and perhaps that he likes potato chips (preferences) [1]. Much like how scientific theories allow scientists to produce a coherent description of natural phenomena, these intuitive theories allow humans to reason about and explain both the physical and the social world that they live in [28], [79].

In recent work, we [58], [59] and others [68], [85] have proposed that people also possess a structured intuitive theory of emotions. This intuitive theory comprises conceptual knowledge such as: what are emotions, moods and other affective states; what types of event outcomes cause emotions; what mental states influence emotions; and what types of behavior emotions influence. Importantly, this conceptual knowledge goes beyond encoding simple pairwise emotion-behavior contingencies, such as recognizing emotional facial expressions, which many machine learning models are already adept at [90]. It allows reasoning about emotions in more complicated situations, such as with novel events or in a new (e.g., cultural) context, and it enables explanation generation, counterfactual reasoning, and hypothetical reasoning [59]. Laypeople use their rich intuitive theories effortlessly in daily life—the challenge for researchers is to distill this implicit knowledge into workable computational models.

We note that there are important differences between scientific theories of emotion [17], [60] and the intuitive theories that laypeople employ. For one, the lay description of emotion is far more coarse-grained. Intuitive theories do not care about the neural bases of emotion [45], nor are they concerned with a taxonomy of how many emotions there are, or whether there is a hierarchy between “basic” and other emotions [15], [61]. In fact, these intuitive theories differ remarkably across cultures—the antecedents and consequences of emotions like shame vary by culture [81], [83]—as well as across individuals [57], based on relationship history [55], mood [11], or even mental illness [12]. In other words, while scientific theories seek an “objective” truth, lay theories describe the subjective reality that people live in. Thus, intuitive theories are crucial targets for computational modeling because they are what laypeople use to make sense of the emotions of others.

The intuitive theory approach lends itself well to computational modeling, especially probabilistic graphical modeling. A probabilistic graphical model represents random variables as well as the probabilistic interdependencies between these variables [40]. For example, a widely supported intuitive theory of human behavior holds that (people think that) other people have beliefs about the state of the world, have desires—goals that they want to achieve—and subsequently form intentions to act upon their beliefs to maximize their desires [10], [46]. This belief-desire psychology, or belief-desire-intention psychology, can be formalized as a rational agent acting to maximize its utility given its incomplete knowledge about the world, and implemented using decision networks or partially-observable Markov decision processes [2], [34], [35]. Inferring beliefs and desires from actions reduces to Bayesian inference over this generative model [1]. These recent “Bayesian Theory of Mind” models have been successful at predicting participants’ judgments of an agent’s beliefs and desires, suggesting that they are capturing how people reason about others’ goal-directed behavior.

We can also add emotions into these generative models of behavior. People know that others’ emotions arise as a reaction to a motivationally salient event in the world, based upon their beliefs and desires about the world [69], [78]. Many scientific theories of emotion propose that people implicitly evaluate (“appraise”) experienced events along a number of self-relevant dimensions based upon their mental state, and feel emotions as a consequence of this appraisal [17], [60], [72]. For example, a goal-conducive event like receiving tenure would produce happiness, while an unexpected event would give rise to surprise. Recent work has also suggested that laypeople perform a similar, “third-person appraisal” process to reason about the emotions of others [8], [58], [70], [78], [82], [85]. When reasoning about someone else’s emotions, people implicitly adopt that individual’s perspective and appraise the situation on behalf of that individual. This capability develops very early in life: even infants and young children are able to reason about the emotional expressions that accompany fulfilled or thwarted goals [71], [80] or mismatched expectations [56], [86]. Thus, a simple intuitive theory of emotions might have event outcomes and mental states jointly “causing” emotions via an appraisal process, which in turn “cause” facial expressions and other behavior [8], [58], [68], [85].

Several recent studies have modeled such intuitive reasoning about emotions using probabilistic graphical models (Fig. 1). We [58] investigated the third-person appraisal process when people evaluated the emotions of others playing gambles. We also had participants reason about emotions given both the gamble outcome and another behavioral cue (e.g., facial expression or verbal utterance). Such complex inferences from multiple cues—cue integration—can be mathematically derived using Bayesian inference [89], and the predictions from a Bayesian model track laypeople’s judgments in these complex scenarios. De Melo and colleagues [8] studied laypeople’s inferences...
of appraisals from emotional expressions by modeling the generative processes from outcomes to appraisals, emotions and expressions. Given an emotional expression, one can then compute an estimate of the latent appraisals by doing reverse inference in the model, a process they term "reverse appraisal". Indeed, they find that laypeople’s subsequent behavior towards the agent are mediated by their inferred appraisals of that agent’s emotions. Wu and colleagues [85] similarly proposed a model of how mental states—beliefs and desires—give rise to actions and emotions; They then model inference of mental states from emotional expressions via Bayesian inference, and show that these also track participants’ judgments. Although these independently-conducted studies contain differences in the details of their models (Fig. 1), there was surprising agreement in the broader approach of using Bayesian inference on a generative model of emotion, and even in the general causal flow of the model. For example, each study highlighted the importance of the third-person appraisal process as an antecedent to emotion (for a more in-depth discussion of these approaches, we invite readers to see [59]).

From a computational cognitive science standpoint, inference in these probabilistic models of emotions track laypeople’s judgments; this shows that such models provide a good computational account of laypeople’s psychology and offers valuable scientific insight. From an affective computing standpoint, such approaches provide strong, theory-grounded models that can be used to build applications that reason like laypeople, and applications that laypeople will reason about. For example, if the goal is to build an expressive virtual character [73] or affective tutor [7] that interacts with laypeople, then perhaps an intuitive theory-based model may provide a way to generate more human-like affective behavior.

Implementing intuitive theories as probabilistic graphical models has its limitations. In our opinion, the most significant limitation is representation. Variables like emotion are often represented using a single real-valued number, or a vector of numbers along a number of emotion dimensions (see also [87]). These representations cannot easily encode more complex information like the relational and temporal nature of emotions: John is not just angry, 6 out of 7, he is often angry at something or someone; this anger may be fleeting, or may give rise to a life-long resentment. Such information is crucial for proper reasoning about what behavior John might next exhibit or how one might intervene to help regulate John’s emotions, but is difficult to represent using graphical models. Additional representational challenges arise when considering other affective phenomena, such as moods and appraisal. Consider appraisal: Within a constrained scenario such as playing a gamble or a social dilemma game with another agent, the appraisal process may reduce to some linear combination of real-valued features that define the outcome of the game. More generally, however, the appraisal process is a complex evaluative process that may be difficult to represent with current graphical approaches. Finally, learning probabilistic graphical models from data and doing inference on large models is computationally intensive, which limits their real-world applications on larger datasets, especially with naturalistic data. We propose that a modern "successor" of probabilistic graphical modeling—probabilistic programming—may offer a solution to these limitations.

3 IMPLEMENTING INTUITIVE THEORIES USING PROBABILISTIC PROGRAMMING

Probabilistic programming is a powerful probabilistic modeling paradigm that offers much promise for affective computing. Like probabilistic graphical modeling, probabilistic programming allows one to capture abstract, conceptual knowledge (e.g., in human intuitive theories) as generative models. Instead of a graphical representation, probabilistic programming represents such conceptual knowledge as stochastic programs—chunks of code that embed randomness into their execution [24], [27]. We start by introducing two recent examples of probabilistic programming applied to model human cognition (Fig. 2), before discussing its features and its potential application to affective computing.

In a landmark study, Lake, Salakhutdinov, and Tenenbaum [41] introduced a handwriting-recognition model that successfully learns character concepts. After being shown just one example of a novel, handwritten character, the model was able to correctly identify more examples at a level comparable to humans (one-shot classification); and when made to generate new exemplars of a novel character, produced examples that were indistinguishable from those produced by human volunteers.
(human-like generative capacity). Underlying these impressive capabilities is a powerful idea: a probabilistic generative model that models the actual writing process. The model has, as primitives, handwritten strokes like straight lines and curves. These strokes are composed to form more complex parts that in turn are composed to make characters, subject to constraints on where and how different strokes may be joined together to form different parts or characters. Finally, the model allows for motor variance at different steps in the handwriting process, such as in choosing the start point of each stroke, trajectory of each stroke, and how different strokes are joined together. Such variance in the generation process introduces “noise” into the visual appearance of the character, but do not change the underlying concept that is to be conveyed—a badly-written “g” is still an “g”, despite its potential visual similarity to “q” (Fig. 2a). When presented with novel characters, the model infers the steps needed to generate the input characters: it inductively learns a program for producing the character. This allows it to learn abstract conceptual knowledge from visual features, and to flexibly apply this knowledge in a more human-like manner [42].

Probabilistic programming can also be applied to social reasoning, such as language understanding [25]. In conversation, people naturally assume, by convention, that others are being informative when speaking, and this allows people to make pragmatic inferences over what was not explicitly said [29]. A statement like “Some of the students passed the exam” invites the interpretation “Not all of the students passed the exam”, even though the semantics of the latter statement is stricter than and not necessarily implied by the former [62]. We and our colleagues have formalized this reasoning in terms of a Rational Speech Act (RSA) framework for language understanding [23]. First, we define the literal listener (Lit) as a probabilistic program that maps a heard utterance to its possible literal meanings (e.g., “some” denotes a non-zero quantity, which could be one, or two, etc.). Next, the speaker (S) is a probabilistic program that models a cooperative human communicator: It aims to achieve the goal of Lit correctly inferring the state of the world, and chooses an utterance to achieve that goal. Finally, the pragmatic listener (L) is a probabilistic program that hears an utterance and reasons about the goals of the speaker S who produced it. Thus, we implement the pragmatic listener L as a probabilistic program that has within it a nested program S, which has in turn another nested program Lit (Fig. 2b). This allows nested social reasoning: the listener reasoning about the speaker reasoning about the listener. For example, upon hearing “Some of the students passed the exam”, L reasons that, if it were indeed the case that “all the students passed the exam”, then S would have said so; since S could have but did not, it is more likely that “Some, but not all of the students” achieved a passing grade. Note that S and Lit are not actual agents, but they exist within listener L’s intuitive theory of communication, which makes assumptions that speakers choose their behavior rationally and following Gricean maxims. The RSA framework has been applied to model understanding of generic language [75]; nonliteral language like hyperbole [37]; humor in wordplay [36]; and politeness in indirect speech [88]. More generally, using probabilistic programs in a compositional—in this case, nested—manner provides a computational framework for modeling social reasoning.

The two examples above illustrate some of the features of probabilistic programming as a modeling paradigm that make it appealing to affective computing. The core idea is representing theory (laypeople’s intuitive theories about handwriting, communication, or emotions) in terms of probabilistic programs. Unlike deterministic programs which always produce the same output when given the same input, probabilistic programs instead produce samples from a distribution of possible outputs. This allows explicit modeling of uncertainty, whether such uncertainty arises from (i) incomplete knowledge about the world and others’ unobservable mental states, (ii) incomplete theory, or (iii) inherent randomness in the generative process. For example, Lake and colleagues [41] introduce one form of uncertainty via motor variance at various steps in the character generation process, while in the RSA framework [23], the model explicitly represents uncertainty in semantic meaning and speaker goals. Applications of affective computing, in particular, face many sources of uncertainty. Third-person appraisal requires an uncertain inference about others’ latent beliefs and desires (incomplete knowledge). There are individual differences, such as in personality or cultural background, in how people with the same expectations and goals appraise the same outcomes, and how they will behave after. Many of these individual differences have not been explored by

![Fig. 2. Simplified illustration of probabilistic programming applied to human cognition, with pseudocode. (a) Our adaptation of the key ideas in [41]. Characters (e.g., “g” and “q”) are composed of a sequence of parts and subparts and their spatial relations (i.e., where the parts are joined). There is motor variance in the “writing” process, which may lead to visually similar, but conceptually different, character concepts. (b) Language understanding in the Rational Speech Acts framework [23] proceeds via nested reasoning: the pragmatic listener reasons about a nested speaker, which in turn reasons about a nested literal listener.](image-url)
scientists (incomplete theory). Finally, the same person facing the same situation in the same context might not always behave in the same way (inherent randomness). Explicitly modelling these sources of uncertainty is important for learning from data, and generalizing to new agents or new contexts.

A second crucial feature is modularity. In a probabilistic programming paradigm, small, modular probabilistic programs are composed together in a hierarchical and/or sequential fashion to produce more complex phenomena—or to capture reasoning about such phenomena. In the handwriting recognition example, there is a hierarchy where strokes compose to form sub-parts, parts, and characters, while in the language understanding example, social reasoning proceeds via inference in nested programs. Modularity and compositionality are particularly important in modeling emotions and the processes that give rise to and arise from emotion [50], [63]. This is because modeling emotion necessitates modeling a wide range of processes from emotion elicitation (via appraisal) to the behavioral effects of emotion. Modularity is essential for defining small, re-usable processes, and compositionality allows one to structure complex reasoning over these processes. For example, an agent experiencing a negative emotion (after appraisal) may want to down-regulate their negative emotion. One way that it can do that is to choose an action that will likely result in positive appraisals. Thus, the agent has to re-use its concept of appraisal to best select an action to achieve its goals. In fact, this action selection is an example of an inference that is handled naturally in a probabilistic program. With a generative model of actions to outcomes to appraisals to emotions, one can condition on a desired emotion, infer desired appraisals and outcomes, and infer the actions that one has to take to achieve the desired goal.

Probabilistic programming is also becoming easier to implement. There has been a surge of development in probabilistic programming infrastructure in the past decade, due in part to a large funding initiative from DARPA [65]. There now exists many probabilistic programming languages that are under active development, and many of these exist as modules or libraries written in existing programming languages like Python (e.g., Edward [76], Pyro) and Javascript (e.g., WebPPL [26]). These languages are Turing complete, and can represent any computable probability distribution. Moreover, some of these language leverage other existing libraries for optimized deep learning: Edward is built on top of Google’s Tensorflow deep learning library, while Uber’s Pyro is built on top of PyTorch. This allows them to leverage efficient optimization (e.g., gradient descent algorithms) and inference techniques (e.g., MCMC, variational inference), as well as hardware acceleration (e.g., GPU computation). Thus, these modern probabilistic programming languages combine both the modeling flexibility of a universal programming language with the power of modern deep learning.

4 Modeling Emotions using Probabilistic Programs

At an abstract level, probabilistic programming is an approach to modeling: It is not bound to any particular language or even a particular architecture, unlike work in cognitive and emotion architectures [47], [49]. That said, to illustrate our points more concretely and to jumpstart the community’s adoption of a probabilistic programming approach to affective computing, in this section we provide accompanying code written in the open-source probabilistic programming language Pyro, itself written in Python. A repository for the code in this paper, accompanying documentation, and links to tutorials, are available at https://github.com/desmond-ong/pp1AffComp

4.1 Description of dataset

As an illustrative example throughout this section, we use data that we previously collected ([58], Experiment 3; available at github.com/desmond-ong/affCog). In this experiment, participants were shown characters playing gambles for money. On some trials, participants saw the outcome of the gamble, or a facial expression ostensibly made by the character after seeing the gamble, or both the outcome and the facial expression. They then rated how the character felt on 8 emotions (e.g., happy, sad, anger), each using 9 point Likert scales. The facial expressions were all generated using FaceGen, with standardized gender, race, and other features, and varying only in emotional expressions. We previously [58] used a probabilistic graphical modelling approach to show that participants’ judgments of emotions given multiple cues—what we term emotional cue integration—can be modeled as the joint Bayesian inference of P(emotion | outcome, face), using the individual likelihoods P(emotion | outcome) and P(face | emotion).

In this paper, we use probabilistic programming to re-model this dataset. The purpose of this re-modeling is pedagogical. In particular, we use a manageable and real dataset (rather than a simulated, toy dataset) and with actual hypotheses. Our aim is not to outperform our prior analysis or any existing benchmarks, but to prioritize readability and abstract understanding over performance. However, we show later that the probabilistic programing re-modeling offers more capabilities than our prior analysis, such as the ability to generate novel faces.

4.2 Modeling Appraisal

First, let us tackle modeling what “causes” emotion: the appraisal process. Appraisal can be represented by a function that takes in some representation of the event and performs some computation to yield emotions. The model could also define an intermediate representational space of appraisal dimensions, such as goal-conduciveness, novelty, and controllability [4], [60] map events onto those dimensions, and subsequently map from the appraisal dimensions to emotions.

Following our previous work [58], let us consider a simple appraisal model (Fig. 3). People observe an agent experiencing the outcome of a gamble, and provide emo-
tion ratings of how that character is feeling. To do so, we assume that people compute an appraisal of the outcome, which could be a linear function of the features of the gamble. We can, based on theory, identify a set of features of the gamble that may factor into such an appraisal, such as the amount won relative to the expected value of the gamble. Let us abstract out that computation into a compute_appraisal() function that returns a set of appraisal variables. Given these appraisal variables, we can then learn a (linear) mapping to the emotion—really, a linear regression.

This linear model provides a simple starting point that already contains many rich features. We abstracted away the appraisal calculation in a separate compute_appraisal() function, which we can modify without impacting the logic of the generative model. In this example, we modeled the mapping from appraisal variables to emotion ratings using a linear regression; we could decide to substitute that with a non-linear function, like a feed-forward neural network that we can fit to the data. To reiterate, this choice of fitting and learning is conceptually distinct from the choice about the dimensions of appraisal.

Another way of modeling uncertainty, as taken in Bayesian regression, is to model the prior distributions over the parameters of the model. This is useful in modeling individual or group differences in appraisal. For example, people reason very differently about the emotions of someone close to them than a stranger, and we previously showed that this difference could be captured via differences in model parameters [57]. In Bayesian regression, one can hypothesize that there are several populations of people, each with their own distribution of model parameters—perhaps males tend to have different parameters from females. We posit prior distributions over parameters, and we can sample regression models and their parameters from those distributions. Conditioning on the data, we can infer population distribution of model parameters, which we then interpret to draw conclusions about group or individual differences in emotion reasoning (We provide an example of this in our repository).

Representing appraisal as a probabilistic program allows many exciting extensions. For example, the appraisal function can also take as input an estimate of the agent’s mental states like beliefs and desires. Depending on the model specification, one could first infer the beliefs and desires, and then pass them into an appraisal function to reason about the latent emotions [8], or one could jointly infer beliefs, desires and appraisals conditioned on the data [85]. We can also think about defining a richer representation for emotions, and doing computation on that representation. In the example above, the return value of the compute_appraisal() function is a real-valued vector (or tensor) representing emotion intensity: This function could instead return an emotion “object” which might contain attributes such as the target of the emotion. To achieve this, we would have to define a space of possible targets (which could be constrained given the specific context to the agents present) so that we can sample from this space during inference, and we would have to embed the target information within the appraisal representation. Although implementing this would still require effort, it seems to us that probabilistic programming offers the most plausible route to success, compared to existing approaches.

4.3 Learning emotion recognition from faces

Next, we consider how we can easily integrate learning from high-dimensional data such as raw images. For most affective computing applications, we are interested in the mapping from emotions to observable behavior like facial expressions. While there are theories that map emotions to facial expression (e.g., the Emotion-Facial Action Coding System or EM-FACS; modified from [16]), they may not be exhaustive, and implementing them to work on raw pixels brings its own set of engineering challenges [6]. A machine learning approach would learn the mapping from high dimensional facial expression images to emotions, rather than specify the mapping by hand.

We can extend the code above to specify a generative model from emotions to facial expressions, \( P(\text{face} \mid \text{emotion}) \), and to learn the parameters \( \theta \) of such a model from data. For example, we can use a convolutional neural network to model \( P(\text{face} \mid \text{emotion}, z) \), where \( z \) is a latent vector that captures aspects of the face that are not determined by the emotion (e.g., face shape, gender). The parameters \( \theta \) can be learnt via stochastic variational inference (SVI) [32]. Modern probabilistic programming languages are able to perform SVI automatically with a small amount of input from the modeler. SVI historically re-
quired the derivation of a quantity called the evidence lower bound (ELBO)—the ELBO is maximized during training, much like how a loss function is minimized during many machine learning approaches. In practice, the ELBO contains the posterior distribution (P(z | face)), which is often intractable, but can be approximated with variational distributions (for example, some q(z | face)) that can be parameterized by neural networks. Trained in this manner, the model is a semi-supervised variant of the variational autoencoder (VAE) [39], a popular generative model that has received significant attention in the deep learning community.

Building this model in Pyro is relatively straightforward (Fig. 4), and again highlights how the complexities of inference and optimization are orthogonal to model specification. In the model, using a logic similar to Fig. 3, we simply have to sample the z’s from some priors (that could also be learnt). We then specify a Decoder() function as a neural network that takes the z’s and the emotions and generates a corresponding face. Within PyTorch, a fairly complicated neural network can be specified in just a few lines of code (see our repository).

This model can be further extended by letting the model learn the latent emotion space. Thus far, we have treated emotion as an observable variable, as we operationalized emotions using ratings that people provided, either after an agent experienced an outcome, or after an agent shows an emotional facial expression. This is usually standard practice in many supervised learning paradigms (e.g., see reviews of [90]), where we train a model to predict an observed y given variables x and provide the model with fully-observed pairs of (x,y). However, in operationalizing emotion as a set of Likert scale ratings (as we have done), or a categorical label (as many classification tasks do), we limit the model to learn only the emotions that we, the modelers, chose, and only in the manner we specify. Concretely, some emotion theories have argued that there is a latent affect space that characterizes affective phenomena, and that most of the variance in emotion ratings or emotion concepts can be captured by a low-dimensional representation, such as two-dimensional space containing Valence and Arousal [67], or three dimensions or Valence, Arousal, and Dominance [52].

To build such a model, we adapt the structure of a recently-proposed Multimodal Variational Autoencoder [84]. In this variant of the VAE, given in Fig. 5, we posit a latent affect space. Outcomes, via appraisal, give rise to changes in the agent’s affect, which in turn give rise to the agent’s facial expressions. When participants report on the agent’s affect by providing emotion ratings, they are mapping their estimate of the agent’s affect onto lay emotion concepts via some mapping P(rating | affect). Note that the rating space has dimensions of discrete emotions like happiness, sadness (i.e., the dimensions of the Likert scales that we asked participants to fill out), but the affect space is one that we allow the model to learn from the data. The model might also learn latent features about the faces, outcomes, or ratings that are unrelated to emotions,
such as gender, face shape, much like the latent variable $z$ that we added into the model in Figure 4.

The multimodal VAE model provides some flexibility in performing inference. Because of the causal dependencies in the graphical model (Fig. 5, right), each of the modalities are mutually independent given the latent affect. This means that the model can deal effectively with any subset of the modalities. For example, it can do inference about emotion ratings if given examples with only facial expressions, only outcomes, or examples with both facial expressions and outcomes; it essentially "ignores" any contribution from missing modalities. This is a big advantage over many existing multimodal emotion recognition systems that cannot flexibly deal with incomplete data (i.e., observations with missing modalities).

### 4.4 Using probabilistic programs as generative models

All the models discussed here are generative models that allow us to sample new examples from the model: indeed, sampling and conditioning is part of the inference process. In particular, if we consider the multimodal VAE model from Fig. 5, we can randomly sample a value of affect in the latent space (i.e., we can sample from the prior, often a Gaussian sphere), and generate the outcomes, faces, and emotion ratings that are associated with that value of "affect". We can also perform conditional sampling: we can condition the model on a particular set of emotion ratings (e.g., high on happiness, high on surprise), and have the model generate the outcomes and faces that are most likely associated with those ratings. In Fig. 6, we show some conditional samples from the model. Given that we only showed the model a small set of faces, it was able to reproduce the faces fairly well, and in fact it can generate new faces that correspond to a novel outcome or an uncommon set of ratings.

In these examples, we have avoided discussing inference in these models for a number of reasons. First, we chose to focus on motivating the model specification from emotion theory. Second, Pyro—and PyTorch—abstracts away the optimization and inference algorithms from the affective computing modeler using high-level functions. Indeed, there exist powerful routines in Pyro like stochastic variational inference [32] that solve or approximately solve the types of problems that arise in affective computing. Third, we show that the modeler can leverage state-of-the-art research done in artificial intelligence and deep learning, such as easily adapting the semi-supervised and multimodal [84] variants of the variational autoencoder [39]. This can be done without necessarily getting bogged down by the details of how and why they work. For readers interested in the details of the implementation, we provide more details about inference, and links to literature and tutorials, in our code repository.

### 5 General Discussion

In this paper, we propose that probabilistic programming offers a principled and theory-driven, yet flexible and computationally-efficient manner of specifying affective computing models. We draw inspiration from recent lay theories of mind and lay theories of emotion implemented using probabilistic graphical approaches, as well as two recent examples of non-affective models implemented in probabilistic programming, to propose a basic framework for modeling emotions using probabilistic programs. We provided illustrative code, written for conceptual clarity rather than predictive performance, about how one might model appraisal (reasoning about emotions from outcomes that occurred), emotion recognition from images, and inference about emotion from multiple channels.

We hope that, by providing open-source code in a modern probabilistic programming language, we might lower the barrier to entry for two groups of researchers. The first includes emotion theorists who want a standardized approach to specifying computational models of emotion, which is both easy to learn yet flexible enough to represent complex theory. For this first group, being able to leverage optimization and deep learning libraries to learn efficiently from large data is an added bonus. The second group includes machine learning researchers and computer scientists familiar with deep learning who want to build more psychologically-grounded models, by offering a modeling paradigm that elegantly specifies scientific theory. This will hopefully provide a bridge to synergize efforts in affective computing from both data-driven and theory-driven approaches.

### 5.1 Boundaries of the probabilistic programming approach

The probabilistic programming approach is suitable to implement a wide class of models. One way to characterize this class of models is by borrowing David Marr’s [48] classic proposal that researchers can understand computation (and cognition) at three complementary "levels of analyses". The highest-level, the "computational level", Fig. 6. Samples from a multimodal VAE. Top: sampled faces conditional on very low and high reward outcomes. Bottom: sampled faces and gamble outcomes, conditioned on emotion ratings. The black triangle on each wheel indicates the outcome of that gamble.
focuses on understanding the goal of the computation: what are the inputs needed to produce the output of a computation. For emotion recognition, what are the contextual cues and behaviour needed for inferring someone’s emotions? Probabilistic approaches, as they focus on building causal models of the world, are well-suited to models at this level of analysis: Indeed, all of the models discussed in this paper are framed at this level.

By contrast, probabilistic approaches tend to be more agnostic about Marr’s second, “algorithmic” level of analysis. Models at this level are concerned with the process of transforming the inputs of the computation into the output. Questions at this level include: how do people scan facial expressions for emotional information—eyes, then mouth? How fast do people make these judgments? How long do emotion episodes last? At the present moment, most probabilistic approaches, including probabilistic programming, do not make strong commitments to process-level models, or resultant behaviour like eye-tracking and reaction times. It remains to be seen whether future probabilistic approaches can make stronger claims at this level of analysis. This is in contrast with, for example, cognitive and emotion architectures [13], [47], [49], which usually take a strong theoretical stance on the dynamics of emotion processes, such as defining fixed-interval cognitive processing cycles and similar constraints on emotional dynamics. (To complete the discussion, probabilistic approaches have even less to say about Marr’s third, “implementation” level, which is concerned with how computation is implemented in the brain or other physical systems. No one of Marr’s “levels” is superior to the others: They answer different questions, and are all complementary to understanding computation.)

5.2 Connection to other approaches

In Sections 2 and 3, we discussed how the probabilistic programming approach can naturally implement intuitive theories (or similar, probabilistic graphical approaches). It is also compatible with many existing non-probabilistic models. For example, Ortony, Clore and Collins’ [60] model of appraisal is written as logical rules, and can easily be modified and implemented in a probabilistic program. Probabilistic programs can also implement dimensional representations of emotions and affect [52], [67] or other types of representations [87], and be used to compare these different theories.

As mentioned in Section 5.1, probabilistic programming approaches focus on a different “level of analysis” as process-level models and architectures that describe emotion dynamics [13], [47], [49]. Indeed, we think that in the future, models built using the probabilistic programming approach will have to interface with architectural models when it comes down to defining dynamics. This should be possible given the modularity of both the architectural approach and probabilistic programming. This might also allow the modelling of more task-general aspects of cognition and emotion using probabilistic programming in the future.

Probabilistic programming can also leverage deep learning models and neural network architectures. For example, Fig. 4 and Fig. 5 illustrate variants of a deep generative model. The compositionality of probabilistic programming allows different "components" of the model to be represented using deep networks [5], [30], [33], [90], but embedded within a larger theoretically-grounded model.

5.3 Future potential

Probabilistic programming may help offer integration across the field of affective computing. Having a common modeling paradigm—and ideally, a common programming language—makes it easy to scientifically test competing theories. Because theories are represented as modular programs, one can easily substitute different theories (via substituting different chunks of code) within the same framework. For example, one can test different formulations of appraisal, or different processes, within this common framework. Many affective computing researchers [50], [63] have repeatedly echoed the need for a common platform to synergize research efforts across the field. While previous cognitive architectures like SOAR had also seemed to promise such a common platform on which to build and test cognitive and emotion theories, the (lack of) uptake of these architectures outside the groups that developed them—relative to the uptake of machine learning and deep learning approaches to emotion—suggests that there may be numerous barriers to adoption for other scientists. Perhaps the solution may lie in having accessible software packages written in popular languages supported by a vibrant and visible development community.

Probabilistic programming may also inform greater insights into understanding human psychology. If we write and train a probabilistic program to model how people recognize emotions from facial expressions, we would be able to learn not only what facial features people tend to use in emotion judgments, but also people’s relative weighting of those features, which might also depend on context or individual differences. But beyond learning people’s “model parameters”, probabilistic programming also allows us to refine existing theories. For example, researchers can specify theories of appraisal as the model’s prior “knowledge”, and the model can update these appraisal mappings to match empirical data, or to even learn new types of appraisals (e.g., using tools from Bayesian non-parametrics). Researchers can then query the programs to further refine their theories of emotion. This is especially necessary today, where there is often too much data for researchers to specify everything necessary for modelling. We need models that can not only implement psychological theory, but also learn to add to and refine existing theories.

As probabilistic programs are generative models, they hold promise for applications which require emotion generation, such as in virtual characters [73]. We showed some simple code that could learn to generate an emotional face conditioned on receiving a particular outcome. Obviously, this model would need a lot more data to generate realistic emotional expressions, but it could already generate emotion-appropriate faces without a programmer telling it what a smile is. Thus, this approach of learning parameters of a generative model from human rat data would be able to augment existing models of emotion generation that rely on hand-tuned expressions.

We can take the idea of emotion generation one step further, to model goal-directed emotion generation. Earlier in
the paper, we discussed the Rational Speech Act framework [23] whereby effective communication can be modeled as nested probabilistic programs. This offers a natural extension to model communicative theories of emotion. Emotions serve a communicative role [43], [78], and people often have goals to convey their emotions to others, for example expressing appropriate negative emotion to an employee who produced sub-par work, or strategically choosing emotions to display in a negotiation [9]. We can apply this principle to model an affective computing agent correctly communicating emotion to a human user. The agent (e.g., a virtual character) builds a model $U$ of how a user inverts the agent’s emotions given the agent’s emotion displays. The agent can nest that model $U$ into its own decision-making model, and chooses its behavior to maximize the probability that the user arrives at the correct conclusion. Thus, by leveraging the compositional nature of probabilistic programs, we can embed communicative theories and goals as nested models of agents reasoning about human users reasoning about agents.

Further extensions may also allow the modelling of complex phenomena like behavioural regulation and deception. For example, people learn to regulate their emotions depending on the social context. An individual inducted at a party may not immediately act upon their anger (to confront their aggressor), and may instead choose to suppress their anger, going against the action tendencies of their current emotional state. They might even go further and fake a smile or a laugh—behaviour contrary to their current emotional state. Current models cannot handle such complex cases, but the combination of probabilistic programming models of emotion understanding discussed here with decision-making models (e.g., POMDPs) may offer a solution. This necessitates several nested layers of reasoning made possible by compositionality: The affective computing agent reasons about how the individual feels in context, but that individual may in turn be thinking about others in their social context and their inferences about said individual.

In summary, affective computing research has made much headway over the past two decades, buoyed by the emergence of many theory-based computational models of emotion, as well as many data-driven machine learning approaches. There still remains, however, much room for integration across many of these research groups and research approaches. It is our hope that the probabilistic programming paradigm, in combining the strengths of both theory-driven and data-driven approaches, may serve as a candidate for a standardized paradigm for unifying efforts in computationally modeling emotion and other affective phenomena.

**Acknowledgment**

We thank Mike Wu and Zong Xuan Tan for discussion and help with coding, and Patricia Chen for feedback on the manuscript. This work was supported in part by the A*STAR Human-Centric Artificial Intelligence Programme (SERC SSF Project No. A1718g0048; Singapore MOE AcRF Tier 1 (No. 251RES1709) to HS; and NIH Grant 1R01MH112560-01 to JZ. This material is based on research sponsored by DARPA under agreement number FA8750-14-2-0009.

**References**


ONG ET AL.: APPLYING PROBABILISTIC PROGRAMMING TO AFFECTIVE COMPUTING


[71] Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. Cognition, 130(2), 204-216.


Desmond C. Ong received his Ph.D. in Psychology and M.Sc. in Computer Science in 2017 from MIT. His research interests include building computational models of emotion and mental state understanding, using a mix of human behavioral experiments and modeling approaches like probabilistic modeling and machine learning. He is a member of the IEEE Computer Society.

Harold Soh received his Ph.D. in EEE from Imperial College London in 2012 where his work on assistive robotics was nominated for the Eryl Cadwaladr Davis Prize and shortlisted for the 2012 UK James Dyson Award. He did his postgraduate work under a SMART Scholar Award at MIT’s SMART Center in Singapore. He is currently an Assistant Professor at the Department of Computer Science at the National University of Singapore. His primary interests are in human-robot/AI collaboration and machine-learning models that enable robots to better cooperate with human teammates. He is a member of the IEEE Computer Society.

Jamil Zaki received his Ph.D. in Psychology from Columbia University in 2010 and did his postdoctoral work at Harvard University. Since 2012, he has been an Assistant Professor of Psychology at Stanford University. He has won numerous awards, such as the 2017 Sage Young Scholar Award, a 2016 Early Career Award from the Society for Social Neuroscience, a 2015 NSF CAREER Award, and a 2015 Janet T. Spence Award for Transformative Early Career Contribution and a 2013 Rising Star award, both from the Association for Psychological Science. His research interests include empathy and emotion understanding.

Noah D. Goodman received his Ph.D. in Mathematics from the University of Texas at Austin, and did his postdoctoral work in the Brain and Cognitive Sciences department at the Massachusetts Institute of Technology. He is an Associate Professor in the Departments of Psychology and Computer Science at Stanford University, and also holds a courtesy appointment in the Department of Linguistics. His work has been recognized by several awards such as the Alfred P Sloan Research Fellow in Neuroscience and a J. S. McDonnell Foundation Scholar. His research interests include language understanding, social reasoning, and concept learning. In addition, he builds enabling technologies for probabilistic modeling, such as probabilistic programming languages.