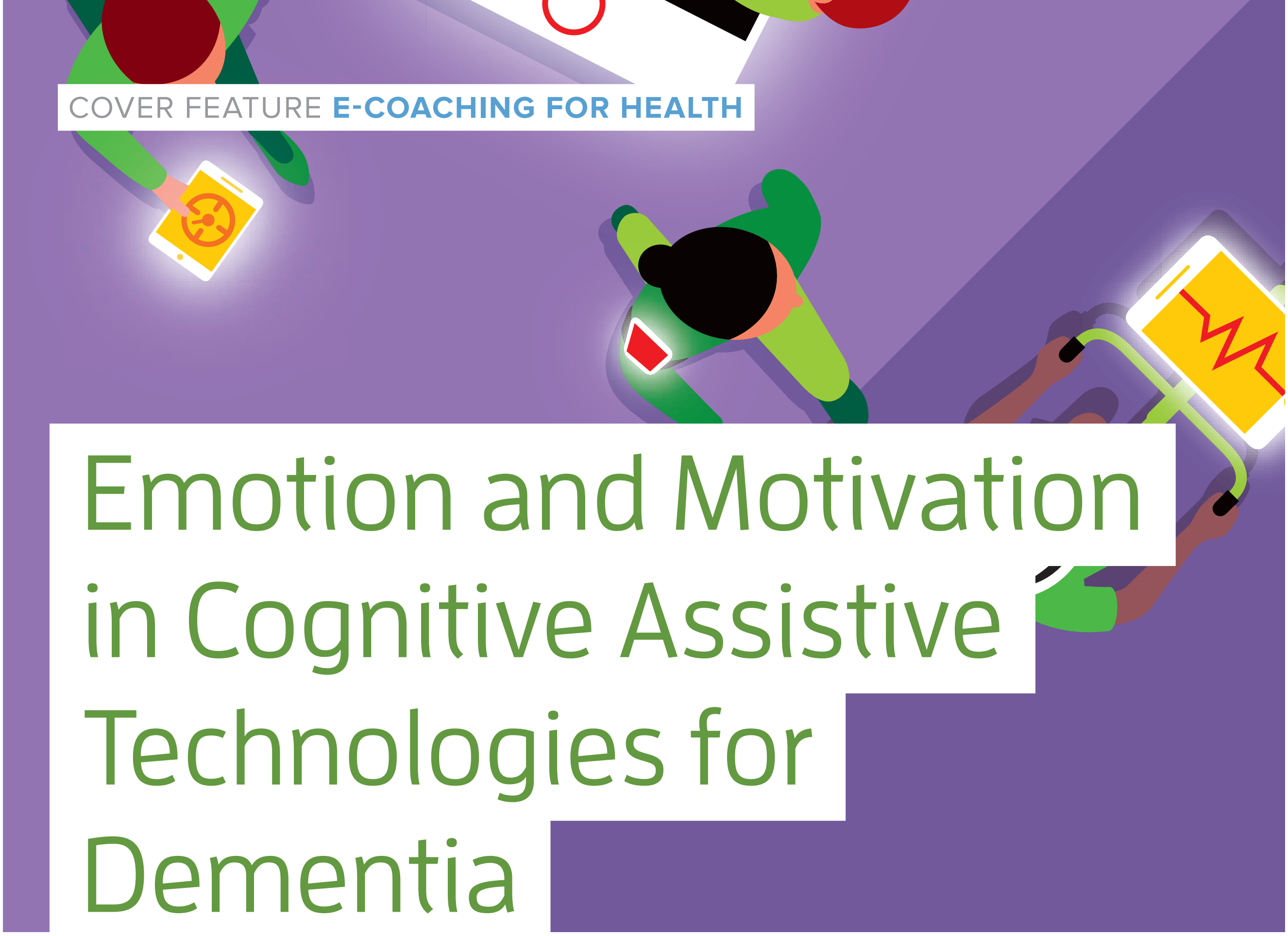

Emotion and Motivation in Cognitive Assistive Technologies for Dementia

Julie M. Robillard, University of British Columbia and BC Children's and Women's Hospital and Health Centres

Jesse Hoey, University of Waterloo

The adoption and effectiveness of cognitive assistive technologies hinge on harnessing the dynamics of human emotion. The authors discuss seminal advances in the integration of emotions in assistive technologies for dementia and propose Bayesian Affect Control Theory (BayesACT), a quantitative social-psychological theory, to model behavior and emotion in such systems.

Technology that persuades people to act must understand and leverage the shared structure and dynamics of human emotion. Intelligent assistive systems for dementia have had limited success in practice, and we hypothesize that integrating emotional reasoning will lead to greater effectiveness, acceptance, and usability. Toward this end, we propose Bayesian Affect Control Theory (BayesACT), a quantitative social-psychological theory about how people perceive others and act socially, to model behavior and emotion in assistive systems and, in turn, promote the alignment of these technologies with users' values and needs. Integrating BayesACT in assistive technologies offers many potential benefits but also presents several challenges.

ASSISTIVE TECHNOLOGY FOR DEMENTIA

Dementia is characterized by the progressive deterioration of cognitive and functional capabilities, leading to a loss of the capacity to perform activities of daily living

0018-9162/18/$33.00 © 2018 IEEE

such as bathing and medication taking. Intelligent assistive technologies have been proposed as a possible solution to support people with dementia in performing these activities independently.[1]

COACH (Cognitive Orthosis for Assisting aCtivities in the Home) is an example of one such technology.[2] It uses a probabilistic and decision-theoretic model of a given daily activity such as hand washing, tooth brushing, or cooking. The model maps from various sensor inputs, including cameras, to a set of prerecorded audio and video cues that are played when the person stops making progress in the task. When tested in a long-term care facility, COACH was found to reduce the need for human assistance, in some cases by as much as 100 percent.[2] However, for some older adults, COACH failed to provide appropriate assistance, leading to confusion or agitation. This result might be due to an emotional misalignment of COACH with the specific needs or personality of the individual user.[3]

Although significant effort has been made to design prompts based on the methods and styles of human caregivers, such as task-focused strategies or paraphrased repetition,[4] a simple one-size-fits-all style of prompting could be limiting. For example, some individuals might respond to a more servile approach, while others might prefer a more imperative style. These responses can be predicted to some extent by models for technology adoption[5] and are influenced by a range of factors such as personal background, sense of self and identity, and emotional responses to prompts, whether given by human or machine.

Recent interviews with older adults with dementia and their caregivers show that emotional processing remains considerably more intact than cognitive processing in dementia, particularly in social situations.[3] These results support the proposition that, while beliefs grounded in cognitive memories fade as people lose their ability to remember people and events, affective aspects could persist longer, even without situational context. Therefore, explicit models of affective meanings offer an attractive mechanism for developing more personalized assistive technologies for dementia.

While Damasio's work is not without criticism, emotional reasoning has emerged as a necessary guide for cognitive deliberation. To some extent, these considerations have been left out of the quest to build artificial intelligence (AI), which to date has largely focused on mapping the perception of stimuli to action by invoking rational utility. Rational AI agents must have preferences that obey certain axioms, and must be able to assess probabilities of outcomes. However, rational decision making engenders numerous paradoxes and inconsistencies,[7] which researchers often attribute to the limited capacity of human intelligence to hold axiomatic preferences or estimate correct probabilities. From this perspective, paradoxes and inconsistencies exist only because of a shortcoming of human intelligence.

Cognitive scientists have attempted to build theories of intelligence that account for deviations from rationality—for example, by following Herbert A. Simon's view of emotions as interrupting mechanisms to cognitive processing so that the system can attend to urgent needs in real time,[8] by modeling the impact of emotion on decision making through the lens of behavioral economics,[9,10] or by focusing on the function of emotion as "bridging the gaps of rationality."[11]

EFFECTIVELY INTEGRATING AFFECTIVE RESPONSES INTO ASSISTIVE TECHNOLOGY REQUIRES A CLEAR, OPERATIONAL DEFINITION OF EMOTION.

INTEGRATING EMOTIONS IN COMPUTING

Effectively integrating affective responses into assistive technology requires a clear, operational definition of *emotion*. Emotions are increasingly recognized as critical components of decision making and human social action.[6] Antonio Damasio's hypothesis that emotions guide behavior contrasts with the Platonic "high-reason" view of intelligence, in which rationality is primarily used to make decisions. Damasio argues that learned neural markers focus attention on likely successful actions, and act as a neural bias allowing people to work with fewer alternatives. These somatic markers are interpreted internally as "cultural prescriptions" for behaviors that are rational relative to social conventions.

Rationality is at the forefront of most of these theories: goals and plans are analyzed and interpreted, and emotions are subsequently generated, preparing the agent for action.

In many of these theories, emotions are viewed as discrete labels describing categorical primitive feelings (for example, "happy" and "sad"), or as a single dimension of valence (good vs. bad). However, this view is at odds with the hundreds of different subtle emotional states humans can describe. Theorists usually propose some mixing mechanism to account for this variety, but it is unclear how this process occurs. This view is challenged by Lisa Feldman Barrett, who argues that there is little empirical evidence to support the idea of discrete emotions like "anger" and "sadness" as fundamental entities in humans.[12] Although some emotion theories do not take the emotions categories to be natural kinds, most still use the primitive emotion categories as endpoints of the analysis. James A. Russell and Barrett propose that a 2D "core affect" might be the fundamental affective entity that is first elicited by stimuli and then categorized using appraisal processes to arrive at interpretations of emotions.[12,13] This core affect forms the basis of "dimensional" theories of emotion. Dimensional emotion has a long history dating back to the 1800s, but modern approaches usually take the view that at least three dimensions are needed (valence, arousal, and potency), possibly four if uncertainty is included.[14] Although popularized in the affective computing literature, the 3D structure of emotions was originally explored by Charles E. Osgood in the 1960s.[15]

ACCOUNTS OF EMOTION IN AFFECTIVE COMPUTING OFTEN CONSIDER THE APPRAISAL PROCESS—WHAT "MAKES" AN EMOTION.

The field of affective computing[16] has yet to fully embrace the deep-seated connection between emotion and action, perhaps because, until recently, there has not been a precise, computationally implementable definition of how emotions guide action.

Accounts of emotion in affective computing often consider the appraisal process—what "makes" an emotion. For example, the classic decision tree in the Ortony, Clore, and Collins (OCC) appraisal model has cognitive appraisal decision nodes and emotions as leaves.[17] Seminal work by Clark Elliott used an OCC model augmented with "love," "hate," and "jealousy" to make predictions about human emotional ratings of semantically ambiguous storylines.[18] OCC models were also integrated with Bayesian networks and probabilistic models for tutoring applications, with a focus on understanding student emotions, but leaving intervention to future work.[19] Klaus Scherer breaks emotion down into five components: appraisal, activation, expression, motivation, and feeling.[20] A behavioral component handles relevance to the organism and prepares possible reactions to stimuli. Emotion is proposed as a facilitator of learning and as a mechanism to signal and predict forthcoming action, but this relationship is not fully elucidated. Other approaches have attempted to reverse-engineer emotion through reinforcement learning (RL) models that interpret the antecedents of emotion as aspects of the learning and decision-making process, but relegate the function of emotion to characteristics of the RL problem.[21] Many of these approaches borrow from behavioral economics and cognitive science to characterize the consequences of emotion in decision making and integrate this knowledge as "coping rules" or "affect heuristics" to influence AI agents' behavior.

In affective computing research, coping is usually modeled as a separate mechanism that maps emotions to a set of action filters that can guide or change decisions made cognitively. Jonathan Gratch and Stacy Marsella[22] proposed a five-stage coping process wherein beliefs, desires, plans, and intentions are first formulated, and upon which appraisal frames are computed. Appraisals are then mapped to multiple emotions using an OCC model, and these emotions are aggregated using an overall emotional state, or "mood." Coping strategies next apply a set of rules to handle the emotions either inwardly, by modifying elements of the model such as probabilities and utilities, or outwardly, by modifying plans or intentions. Christine Lisetti and Piotr Gmytrasiewicz defined specific coping mechanisms as "action tendencies," highlighting their importance in guiding actions.[23]

The BayesACT model described in this article is based on the social-psychological Affect Control Theory (ACT), which defines emotions precisely as vectors in a 3D sentiment space used for sharing and interpreting cultural expectations. Sentiments, also 3D vectors, are inextricably attached to cognitive symbolic interpretations of identities and actions such as words and gestures. The model is grounded in a very tight connection between perceptions and actions through an emotional channel defining cultural expectations, compelling action systems to be guided by a fast cultural heuristic. BayesACT is consistent with research demonstrating affect and uncertainty as heuristics for decision making,[10] but goes a step further by providing a more complete and dynamic model of emotional responses in social situations. BayesACT fundamentally changes the affective computing paradigm by explicitly putting the emotional and volitional "horse" before the cognitive and deliberative "cart."

SENTIMENTS, ALSO 3D VECTORS, ARE INEXTRICABLY ATTACHED TO COGNITIVE SYMBOLIC INTERPRETATIONS OF IDENTITIES AND ACTIONS.

LIMITATIONS OF CONVENTIONAL AFFECTIVE COMPUTING

Appraisal models have largely been the focus in emotional AI and affective computing. Appraisals usually start by defining a set of variables, each of which models some basic entity such as novelty, control, and uncertainty. However, assessing these variables has some limitations.

In AI, novelty can be used in two contexts. First, in a "decision theoretic" mechanism, novelty acts as a positive reward for exploration: newly discovered elements have the potential to be beneficial, and optimism under uncertainty is the by-product. Second, in an "expectation violation" mechanism, novelty evokes an emotion that guides future action through some coping mechanism. In assistive technology for dementia, framing novelty as a reward or dimension on which to base action choices can be problematic, as memory loss can make events, objects, or persons appear novel when they are not, making modeling difficult without a precise statement of how novelty maps to action as a function of biographical memory.

Further, control and power are normally evaluated with respect to a reward function, as agents who can get higher rewards independently of other agents are more powerful. This feature might be relevant in e-coaching technology, as users could experience feelings of dependence related to their health condition. However, power-based decision making should consider subtle and shifting forms of power and control. For example, older adults with dementia might often feel or express strong control as they are enacting an identity held at some time in the past, yielding a priori expectations and notions of control that are subsequently updated as a result of inference. Finally, uncertainty is a key consideration in developing emotionally responsive e-coaching applications. While many appraisal theories do not incorporate a formal notion of uncertainty, many affective computing and behavioral economics approaches have embraced it.

Overall, most computational models of affect build on conventional AI reasoning techniques (planning, logic, decision theory, and so on) and might fail to capture how people actually make judgments based on culturally and emotionally defined markers. Dimensional theories of emotion have been explored to bridge this gap, and valence has been proposed as a heuristic to guide action based on somatic markers.[6,10] However, these theories provide incomplete support for determining how an agent should use emotion-related information to react to a user in a culturally sensitive way to increase motivation and provide the needed assistance.

AFFECT CONTROL THEORY

In contrast to the cognitive-rational models traditionally used in AI, ACT proposes that the main drivers of action are differences between established cultural *sentiments* and transient situational feelings or *impressions*.[24] This affective discrepancy—called *affective deflection* in ACT—generates an initial response to a situational event and interacts with cognitive processing to adjust and refine actions to meet the situation's real-time demands.

ACT proposes that humans learn and maintain a set of shared cultural

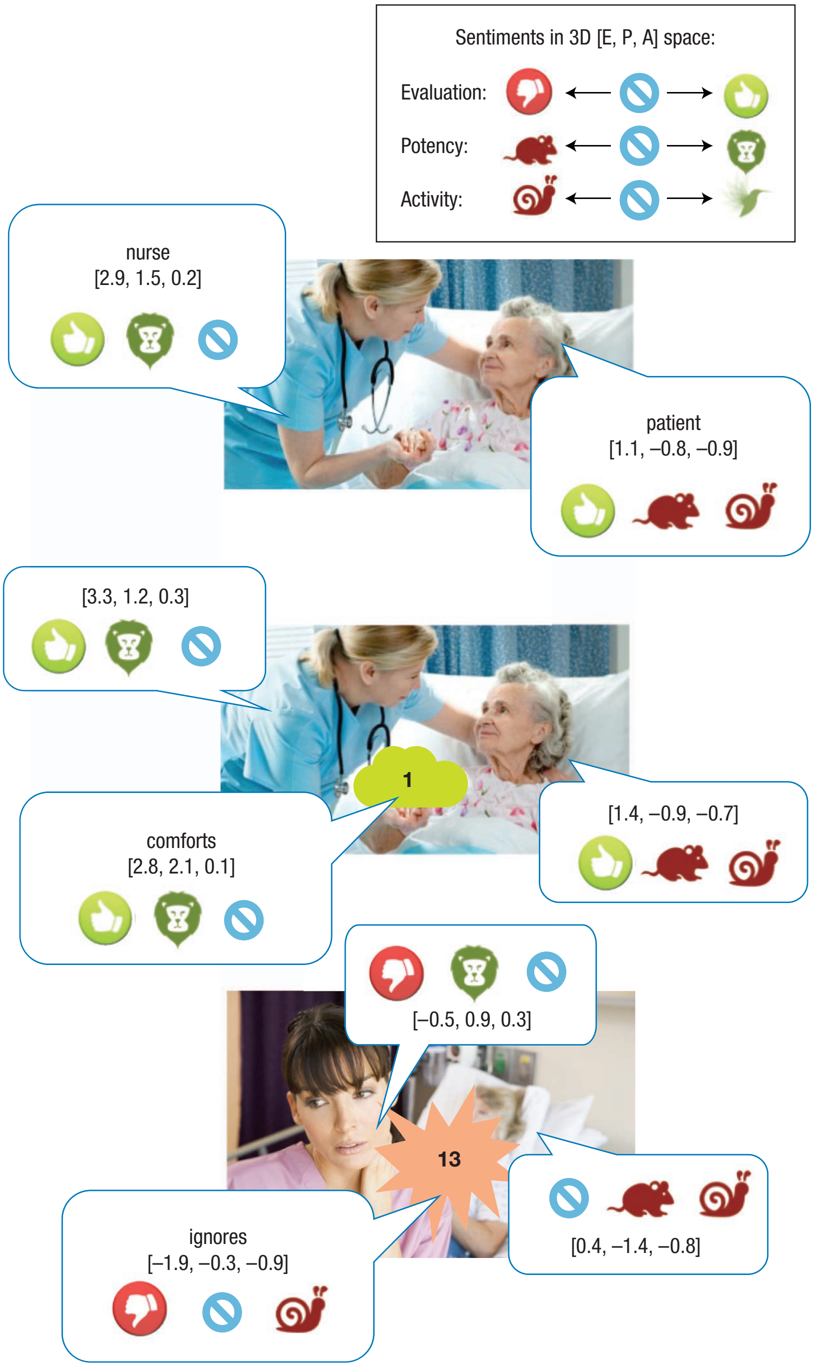

FIGURE 1. Example evaluation, potency, and activity [E, P, A] sentiments about the identities "nurse" and "patient" (top), and the deflections and transient impressions for "nurse comforts patient" (middle) and "nurse ignores patient" (bottom), per the USA 2002–2004 lexicon.

affective sentiments about people, objects, and behaviors, and about the dynamics of interpersonal events. Humans use an affective mapping to appraise individuals, situations, and events as sentiments in a 3D vector space of *evaluation* (E; good vs. bad), *potency* (P; strong vs. weak) and *activity* (A; active vs. inactive). These sentiments can be measured, and their cross-cultural consistency has repeatedly been demonstrated in large studies.[15] Humans use these culturally shared sentiments to make predictions about what others will do and to guide their own behavior, making them a keystone of human intelligence. The shared sentiments, and the resulting affective ecosystem of vector mappings, encode a set of social prescriptions that, if followed by all members of a group, results in an equilibrium or social order. This "affect control principle" has been shown to be a powerful predictor of human behavior. While ACT is based in sociological evidence about human interactions with other humans, people also ascribe affective meaning to media and to technological artifacts.[25]

EPA sentiments can be measured with the semantic differential, a survey technique in which respondents rate affective meanings of concepts on bipolar scales. In general, within-cultural agreement about EPA meanings of social concepts is high, and cultural-average EPA meanings from even a few dozen participants are extremely stable over extended periods of time.[24] Sociologists have gathered EPA ratings for numerous concepts across different cultures (USA 1975/1978/2002–2004/2013–2017, Canada 1980–1986/2001–2003, Ireland 1977, Japan 1989–2002, Germany 1989/2007, China 2001) by surveying thousands of people and compiling ACT lexicons or dictionaries that give average EPA ratings for words for which there is consensus. For example, as Figure 1 shows, the EPA for nurse (from the USA 2002–2004 dictionary, average female rating) is [2.9, 1.5, 0.2], meaning that nurses are seen as very good (E), a bit powerful (P), and a bit active (A). A patient is seen as [1.1, −0.8, −0.9]: comparatively less good, powerful, and active than a nurse. A complete description of these datasets can be found in David R. Heise's *Surveying Cultures*[26] or at research.franklin.uga.edu/act.

Social events cause transient impressions of identities and behaviors that deviate from their corresponding fundamental sentiments. ACT models this formation of impressions from events with a minimalist grammar of the form actor-behavior-object (A, B, O). The ACT predictions are based on an empirically derived impression formation function over the nine-dimension space of $\{E, P, A\} \times \{A, B, O\}$.[24] We denote the fundamental sentiment of the actor's evaluation as A_e, and use A'_e for the corresponding transient impression (and similarly for the other eight combinations). The impression formation function consists of linearly weighted polynomial features that combine fundamental sentiments in ways that represent known psychological consistency effects. For example, the transient impression of an actor's evaluation (A'_e) will be positive if (s)he does something positive (B_e), if (s)he does something either positive to a positive person or negative to a negative person (a *balance* effect, represented by a term $B_e \times O_e$ with a positive coefficient), and if (s)he does something weak to a positive person or powerful to a negative person (represented by a term $B_p \times O_e$ with a negative coefficient).

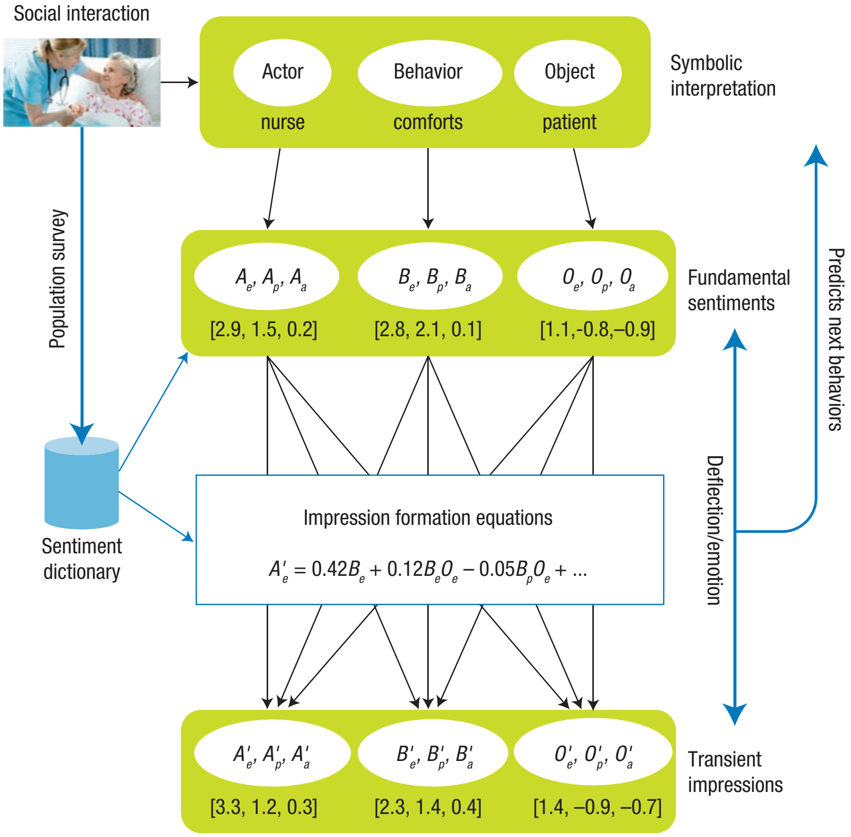

FIGURE 2. Key elements of Affect Control Theory (ACT). An actor (A) performs a behavior (B) on an object (O). A dictionary of empirically measured fundamental sentiments corresponding to the (observed/estimated) behavior and the (known or previously estimated) identities of A and O provides inputs to a set of impression formation equations (also empirically measured). These equations yield transient impressions for A, B, and O. The difference between fundamentals and transients is the deflection, which can be used as a predictor of the most appropriate behavior (for example, a response for the object). The vector difference between fundamentals and transients is the emotion, which is sent as a signal of misalignment.

These three effects are combined with weights showing relative strength as measured in the USA 1978 EPA survey as

$$A'_e = 0.42B_e + 0.12B_eO_e - 0.05B_pO_e + \dots$$

where we only show 3 of the 20 terms in the full equation. The weighted sum of the squared Euclidean distance between fundamentals and transients is the *deflection* and is hypothesized to correspond to an aversive state of mind that humans seek to avoid or minimize. This hypothesis is known as the *affect control principle*.

As a specific example, consider a nurse (actor) who ignores (behavior, with EPA = [–1.9, −0.3, −0.9]) a patient (object). Observers agree, and ACT predicts, that this nurse appears less nice (E) and less powerful (P) than the cultural average of a nurse (transient EPA = [−0.5, 0.9, 0.3]), while the patient also seems less good and less powerful (transient EPA = [0.4, –1.4, −0.8]). The situation in which a nurse ignores a patient has a deflection of over 13 (very high), whereas if the nurse comforts (EPA = [2.8, 2.1, 0.1]) the patient, the deflection is 1 (very low). After comforting, the nurse's transients are very close to her fundamentals (transient EPA = [3.3, 1.2, 0.3]), as are the patient's (transient EPA = [1.4, −0.9, −0.7]). In general, nurses are expected to do positive and powerful things to a patient, and such actions confirm their culturally defined identity sentiments.

The affect control principle allows ACT to predict deflection-minimizing actions for agents by computing derivatives of the impression formation equations. Emotions in ACT are readouts of the 3D vector differences between fundamental and transient sentiments. They are relayed from one agent to another to promote alignment through a set of gestures, vocal tones, and facial expressions. Figure 2 shows a schematic representation of some of ACT's key elements.

A recent generalization of ACT, BayesACT integrates affective dynamics with decision-theoretic reasoning and explicitly models uncertainty as a key element.[27] BayesACT proposes that emotional processing takes place rapidly and continually, while cognitive-rational processing takes any spare cycles to compute optimal plans within the ecosystem of the affective processing unit. BayesACT views feelings or sentiments as bridging the gaps in a social order, while rationality becomes the "interrupt" mechanism that can be used to handle novelty and to repair breakdowns and disruptions.

BayesACT is thus a type of expectation violation model in which we define expectation violation in a 3D sentiment space and use deflection as a term to describe it. Deflection explicitly provides a policy based on identities, which are interpreted as a motivational force. BayesACT is also a decision-theoretic model, since it can compute plans in the (restricted by affect) game tree to optimize over utility, which might include a novelty measure or other appraisals. As deflection grows, people become less certain of themselves, and the breadth of action choices causes a cognitive overload that leads to short-term solutions. These poor short-term solutions could be disruptive for people and their social context. BayesACT resolves this problem with a single parsimonious model of human affect. When a breakdown occurs, for example, after the user gets frustrated, BayesACT can predict the resulting difficulties to make decisions and suggest correct alternatives, or focus the user on one clear identity.

INTEGRATING ACT IN ASSISTIVE TECHNOLOGY FOR DEMENTIA: THE COACH SYSTEM

To demonstrate how ACT can effectively model affective responses in assistive technology for dementia, we integrated BayesACT into the COACH system and used it to explicitly model the identities of the older adult user and the assistive agent. We used these explicit identity models to generate interventions that were tailored to a simulated user's emotional state.

The baseline function of COACH's hand-washing system relies on a Bayesian sequential model with discrete variables corresponding to the different steps of hand washing, describing the state of the tap (on/off) and hands (dirty/soapy/clean and wet/dry). The older adult user's behavior is modeled by his or her actions: turn on/off water, use soap, use towel, rinse, and null (do nothing). There are probabilistic transitions between plan steps described in a probabilistic plan-graph (for example, the user sometimes applies soap first and sometimes turns on the tap first). A binary variable describes whether the user is aware or not, and this is unobserved but inferred from the user's behavior. COACH tracks the user's hands by classifying individual body parts from a single overhead depth image on a per-frame basis. The tracker outputs the locations of the user's two hands and head, followed by a mapping to a set of predefined spatial regions (soap, tap, sink, water, towel), the index of which is used as the observation of the model's state.

The integration of BayesACT requires that COACH be able to learn information about the user's identity. Luyuan Lin and her colleagues experimented with different identities in simulations, including situational identities such as patient and assistant and biographical identities such as boss or athlete.[28] In practice, COACH uses hand coordinates obtained from the hand tracker as a signal of user emotion on the EPA dimensions. However, due to differences in the ability to automatically determine these emotional factors, BayesACT models the strength of the estimates and can factor weak or less informative signals, concordant with the idea of integrating multiple, differently salient sources of information.[28]

Deflection then plays a critical role by modeling how the interaction will deviate from the goal should the system become misaligned with the person. For example, if the deflection is high, then more unusual (for example, nonconforming) reactions by the older adult are expected. Prompting strategies will be more effective if users are aware of their situation and are well aligned with the system, such that each user shares the same emotional models of identity.

While initial tests of the COACH system were promising, more information was required about how to model identities, including about how persons with dementia perceive themselves and a virtual assistant. This knowledge gap was addressed through a set of interviews with older adults, both with and without dementia. The study involved a semi-structured qualitative interview process involving 12 older adult-care residents and 9 caregivers.[3] The interview guide covered life domains (family and origin, occupation/vocation, personal history, and relationships) and feelings related to an intelligent cognitive assistant. All interviews were transcribed and analyzed to extract a set of affective identities and coded according to the social-psychological principles of ACT. The coding was organized around three themes: biographical identities, current identities, and loss of identity or confusion. Once the main identities were extracted, E, P, and A ratings were generated for each identity using standard semantic differential scales.

Thematic analysis of the interviews showed that a set of identities can be extracted for each participant. Furthermore, the results support the proposition that, while identities grounded in denotative memories fade as abilities to remember people and events are lost, affective aspects of identity in the self-sentiment can persist longer.[3] One resident, for example, associated strongly on an emotional level with biographical identities of father (EPA = [1.9, 1.8, 0.0]) and priest (EPA = [2.2, 1.1, −1.5]), even though he was sometimes being treated as a more helpless identity, such as child (EPA = [1.5, −0.8, 2.1]). This person's affective memories did not find situational support, leading to feelings of inferiority.[3]

Future COACH development will include integration of these findings into the system. Beyond COACH, BayesACT simulations have been conducted to model a range of interactions, including to predict older adult responses to online health tools,[29] demonstrating the model's potential applicability across different settings.

BENEFITS OF BAYESACT INTEGRATION

BayesACT integrates three requirements of emotion modeling into a single reasoning framework:

- predicting how people will feel in emotion-evoking situations, building on appraisal models;[22]
- predicting how people will make decisions in emotional states, building on decision-theoretic models;[21] and
- using emotional responses to infer user personality, building on reverse appraisal models and coherence approaches.[30]

Most existing models that attempt to bridge these three requirements focus on task goals and largely ignore social goals.[10] BayesACT not only integrates social goals but also identities, which serve as heuristics for computing plans to achieve goals. Compared with appraisal theories that often consider power as an intermediate construct, BayesACT also harnesses power as a key construct of social interaction. Finally, a practical advantage of BayesACT lies in its dictionaries of cultural sentiments, which act as a large and validated domain theory.

A key barrier to the use of technology, intelligent or otherwise, in applications such as healthcare has been adoption. Despite the creation and development in recent years of assistive healthcare solutions that have been perceived to be useful in theory, and concerted efforts to bring these technologies to market, widespread adoption has been slow. Customizing technology through the integration of BayesACT to increase the likelihood of acceptance and adoption by targeted users might help to shift these technologies from the research realm to the real world. Through the use of BayesACT, technology developers can attempt to predict the affective responses of their users to specific aspects of their forthcoming applications. For example, an AI agent used as a social networking tool or personal assistant might be designed to act in a manner that the user perceives as friendly for those who prefer these types of interactions and more efficient, curt, and distant for those who do not align well with a friendly assistant. Further, BayesACT is culturally sensitive, and dictionaries of affective meanings and impression formation equation parameters have been gathered in many different countries and languages. The consideration of cultural variations in these measurements would result in different behavior from a BayesACT agent.

BayesACT simulations can also be used to optimize user experience with technology to predict the extent to which the technology is likely to achieve trust on the part of the user—a particularly important aspect of a relationship within the healthcare field. As one example, an AI technology could have the aim of aiding its user in online shopping for medication. For users to have positive experiences with such technology, it seems clear that they must, to some extent, trust that the AI agent will make accurate predictions as to what the user wants and needs, and that it will not use that information in any way that the user would disapprove of. Integration of BayesACT into such technology will allow the technology to predict the affective meanings that the user perceives and to then act in a way that is consistent with those meanings, thus promoting trust, engagement, and adoption.

Overall, BayesACT provides an explicit, computational model of identity that allows for the close integration of emotion and action for a wide range of technology applications.

CHALLENGES OF BAYESACT INTEGRATION

Along with promising potential benefits are significant challenges in integrating BayesACT into assistive technology.

A high-level disadvantage is that ACT models are broad; most evidence pertains to how such models predict average group behavior, with less exploration of how they apply to specific individuals in specific situations.

Further, databases of terms and affective meanings take significant time to compile and, given the rapid progress of technology, it might be difficult for database generation to keep pace. For example, the most recent compilation of terms in Ontario, Canada, is from 2003 and has none that relate to technologies as commonplace as iPhones, Facebook, or Twitter. However, to address this issue various techniques can be used to map from known to unknown words. For example, words could be considered "close" based on their semantic similarity ("cat" is closer to "dog" than to "cab"), and this semantic similarity could be learned by observing real-world usage contexts ("the dog chased the cat" but not "the dog chased the vat"). Closeness in emotional space could then also be considered. Alternatively, both emotional and denotative meanings could be simultaneously learned.

A further limitation stems from the fact that to make accurate predictions, BayesACT requires accurate input of the affective meanings that users have toward themselves, the AI agent, and the actions that the agent performs. Incorrect or biased identity inputs could result in the agent performing actions to which the user is entirely unreceptive, which might hinder the trust in and likelihood of continued use of that technology, and, in the healthcare context, could have negative consequences on users' health. Therefore, a promising direction for future research is the development of technology that can actively and nonobtrusively capture the identity of the user of the technology, and efforts in this area are ongoing.

BayesACT integration also raises ethical issues. Technology that considers the affective meanings users hold toward certain objects, persons, and actions, and customizes itself to align with those meanings, could perpetuate objectively negative biases. For example, most existing virtual assistants have, by default, female voices (Apple's Siri, Amazon's Alexa, Microsoft's Cortana). If, for the sake of argument, this is because people associate femininity with subservience, then intentionally using female voices to accommodate the user base clearly perpetuates a negative stereotype. By gauging the affective meanings that a user holds, a BayesACT-infused agent could act in whatever way possible to accommodate the user, even if that means reinforcing particular stereotypes that the user holds.

Finally, among the potential upsides to integrating AI with BayesACT is the ability to foster trust on the part of the user. However, this potential benefit could also be problematic should technology developers and providers use this trust in ethically unacceptable ways—for example, to sway the user into making purchases or engaging in socially irresponsible acts.

LOOKING TO THE FUTURE

BayesACT models human sentiment, rooted in biographical and situational memories of identity, as a core motivational force for human action. It integrates utility and goals, allowing for appraisal dimensions to be explicitly represented if needed. Finally, it provides a clear definition of emotions as signals of incongruence or dissonance in a social interaction. Its probabilistic framework allows for affective information to be quickly used to establish a social context for an interaction, and to thereafter guide action selection in a way that preserves this same context. When interacting with a person with dementia in particular, it is necessary to maintain this social interaction on an emotional level to prevent breakdowns that most often result in a lack of motivation and a lack of action, and thus to undesired outcomes. This necessity is especially critical as older adults with dementia maintain greater and stronger memories on an emotional level than on a cognitive one.

Looking forward, the success of BayesACT integration will rest on how identity is understood. BayesACT operates using databases of identities that are rated based on affective meanings E, P, and A, but it is possible that other theories of identity can illuminate different important factors to be taken into account. Ongoing research is currently examining the predictive power of BayesACT specifically in the context of technological development for a range of different user groups, including online collaborative networks such as GitHub, and workplace settings with persons having intellectual developmental disability. Empirical evaluations of the effectiveness of the model will be followed by large-scale testing and implementation in existing and emerging assistive technologies.

This article has outlined a novel mechanism to allow for culturally shared emotional meanings to be integrated into assistive technology development. These culturally shared meanings guide and motivate human action. Most commonly used appraisal theories of emotions are limited in their consideration of these shared cultural meanings, and therefore have shortcomings that make them of limited utility in the context of applications for people with dementia. BayesACT addresses many of these limitations and clearly defines a way of ensuring that assistive technology is responsive to users on an emotional level. C

ABOUT THE AUTHORS

JULIE M. ROBILLARD is an assistant professor in the Division of Neurology, Department of Medicine, at the University of British Columbia and Scientist in Patient Experience at the BC Children's and Women's Hospital and Health Centres. Her research lies at the intersection of aging, health, and technology, with a current focus on the development of tools for the evaluation of the quality, ethics, and impact of technology for dementia. Robillard holds a BSc in biological sciences from Université de Montréal and a PhD in neuroscience from the University of British Columbia. She is associate director of Neuroethics Canada and sits on the Board of Directors of the Medical Device Development Centre of British Columbia. She is also vice-chair of the Ethical, Legal, Social Impacts Committee of the Canadian Consortium on Neurodegeneration in Aging, a member of the Executive Committee of the Technology and Dementia Professional Interest Area of the Alzheimer's Association International Society to Advance Alzheimer's Research and Treatment, and a Network Investigator of the AGE-WELL Network of Centres of Excellence. Contact her at jrobilla@mail.ubc.ca.

JESSE HOEY is an associate professor in the David R. Cheriton School of Computer Science at the University of Waterloo, where he leads the Computational Health Informatics Laboratory (CHIL), as well as an adjunct scientist at the Toronto Rehabilitation Institute. He works on problems in computational social science, probabilistic and decision-theoretic automated reasoning, affective computing, rehabilitation science, and ubiquitous computing. Much of his research has focused on developing systems to help persons with a cognitive disability such as Alzheimer's disease to engage in daily living activities. Hoey received an MSc in physics and a PhD in computer science from the University of British Columbia. He is a Network Investigator for the AGE-WELL Network of Centers of Excellence. Contact him at jhoey@cs.uwaterloo.ca.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the reviewers and editor for their constructive comments, and Arlo B. Sporn, research assistant at the National Core for Neuroethics at the University of British Columbia, for his contributions to the manuscript, as well as the following funding sources: the Canadian Consortium on Neurodegeneration in Aging, AGE-WELL Inc., The American Alzheimer's Association (ETAC-14-321494), and the Canadian Natural Sciences and Engineering Research Council (NSERC).

REFERENCES

1. E.F. Lopresti, A. Mihailidis, and N. Kirsch, "Assistive Technology for Cognitive Rehabilitation: State of the Art," *Neuropsychological*

Rehabilitation, vol. 14, nos. 1–2, 2004, pp. 5–39.

2. A. Mihailidis et al., "The COACH Prompting System to Assist Older Adults with Dementia through Handwashing: An Efficacy Study," *BMC Geriatrics*, vol. 8, 2008; doi:10.1186/1471-2318-8-28.
3. A. König et al., "Qualitative Study of Affective Identities in Dementia Patients for the Design of Cognitive Assistive Technologies," *J. Rehabilitative and Assistive Technologies Eng.*, vol. 4, 2017; doi:10.1177/2055668316685038.
4. R. Wilson et al., "Quantitative Analysis of Formal Caregivers' Use of Communication Strategies while Assisting Individuals with Moderate and Severe Alzheimer's Disease During Oral Care," *J. Communication Disorders*, vol. 46, no. 3, 2013, pp. 249–263.
5. C. Nugent et al., "TAUT: Technology Adoption and Prediction Tools for Everyday Technologies," *Alzheimer's & Dementia*, vol. 12, no. 7, 2016, pp. P273–P274.
6. A. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, reprint ed., Penguin Books, 2005.
7. D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2013.
8. H.A. Simon, "Motivational and Emotional Controls of Cognition," *Psychological Rev.*, vol. 74, no. 1, 1967, pp. 29–39.
9. G. Loewenstein and J.S. Lerner, "The Role of Affect in Decision Making," R.J. Davidson, K.R. Sherer, and H.H. Goldsmith, eds., *Handbook of Affective Sciences*, Oxford Univ. Press, 2003, pp. 619–642.
10. P. Slovic et al., "The Affect Heuristic," *European J. Operational Research*, vol. 177, no. 3, 2007, pp. 1333–1352.
11. P.N. Johnson-Laird and K. Oatley, "Basic Emotions, Rationality, and Folk Theory," *Cognition and Emotion*, vol. 6, nos. 3–4, 1992, pp. 201–223.
12. L.F. Barrett, "Solving the Emotion Paradox: Categorization and the Experience of Emotion," *Personality and Social Psychology Rev.*, vol. 10, no. 1, 2006, pp. 20–46.
13. J.A. Russell and A. Mehrabian, "Evidence for a Three-Factor Theory of Emotions," *J. Research in Personality*, vol. 11, no. 3, 1977, pp. 273–294.
14. J.R. Fontaine et al., "The World of Emotions Is Not Two-Dimensional," *Psychological Science*, vol. 18, no. 12, 2007, pp. 1050–1057.
15. C.E. Osgood, G.J. Suci, and P. Tannenbaum, *The Measurement of Meaning*, Univ. of Illinois Press, 1967.
16. R.W. Picard, *Affective Computing*, reprint ed., The MIT Press, 2000.
17. A. Ortony, G.L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge Univ. Press, 1990.
18. C. Elliott, "Hunting for the Holy Grail with 'Emotionally Intelligent' Virtual Actors," *ACM SIGART Bull.*, vol. 9, no. 1, 1998, pp. 20–28.
19. C. Conati and H. Maclaren, "Empirically Building and Evaluating a Probabilistic Model of User Affect," *User Modeling and User-Adapted Interaction*, vol. 19, no. 3, 2009, pp. 267–303.
20. K.R. Scherer, A. Schorr, and T. Johnstone, eds., *Appraisal Processes in Emotion: Theory, Methods, Research*, Oxford Univ. Press, 2001.
21. T.M. Moerland, J. Broekens, and C.M. Jonker, "Emotion in Reinforcement Learning Agents and Robots: A Survey," *Machine Learning*, vol. 107, no. 2, 2018, pp. 443–480.
22. J. Gratch and S. Marsella, "A Domain-Independent Framework for Modeling Emotion," *Cognitive Systems Research*, vol. 5, no. 4, 2004, pp. 269–306.
23. C.L. Lisetti and P. Gmytrasiewicz, "Can a Rational Agent Afford to Be Affectless? A Formal Approach," *Applied Artificial Intelligence*, vol. 16, nos. 7–8, 2002, pp. 577–609.
24. D.R. Heise, *Expressive Order: Confirming Sentiments in Social Actions*, Springer, 2007.
25. B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, CSLI Publications, 2003.
26. D.R. Heise, *Surveying Cultures: Discovering Shared Conceptions and Sentiments*, John Wiley & Sons, 2010.
27. J. Hoey, T. Schröder, and A. Alhothali, "Affect Control Processes: Intelligent Affective Interaction Using a Partially Observable Markov Decision Process," *Artificial Intelligence*, vol. 230, 2016, pp. 134–172.
28. L. Lin et al., "Affectively Aligned Cognitive Assistance Using Bayesian Affect Control Theory," L. Pecchia et al., eds., *Ambient Assisted Living and Daily Activities*, Springer, 2014, pp. 279–287.
29. J.M. Robillard et al., "Intelligent and Affectively Aligned Evaluation of Online Health Information for Older Adults," *Proc. 2017 AAAI Joint Workshop Health Intelligence*, 2017; cs.uwaterloo.ca/~jhoey/papers/Robillard-aaai2017.pdf.
30. P. Thagard, "Why Wasn't O.J. Convicted? Emotional Coherence in Legal Inference," *Cognition and Emotion*, vol. 17, no. 3, 2003, pp. 361–383.

See **www.computer.org/computer-multimedia** for multimedia content related to this article.