DOI:10.1145/2950041

Biologically based computational modeling promises virtual characters capable of face-to-face human interaction.

BY MARK SAGAR, MIKE SEYMOUR, AND ANNETTE HENDERSON

Creating Connection with Autonomous Facial Animation

OF ALL THE experiences we have in life, face-to-face interaction fills many of our most meaningful moments. The complex interplay of facial expressions, eye gaze, head movements, and vocalizations in quickly evolving "social interaction loops" has enormous influence on how a situation will unfold. From birth, these interactions are a fundamental element of learning and lay the foundation for successful social and emotional functioning through life.

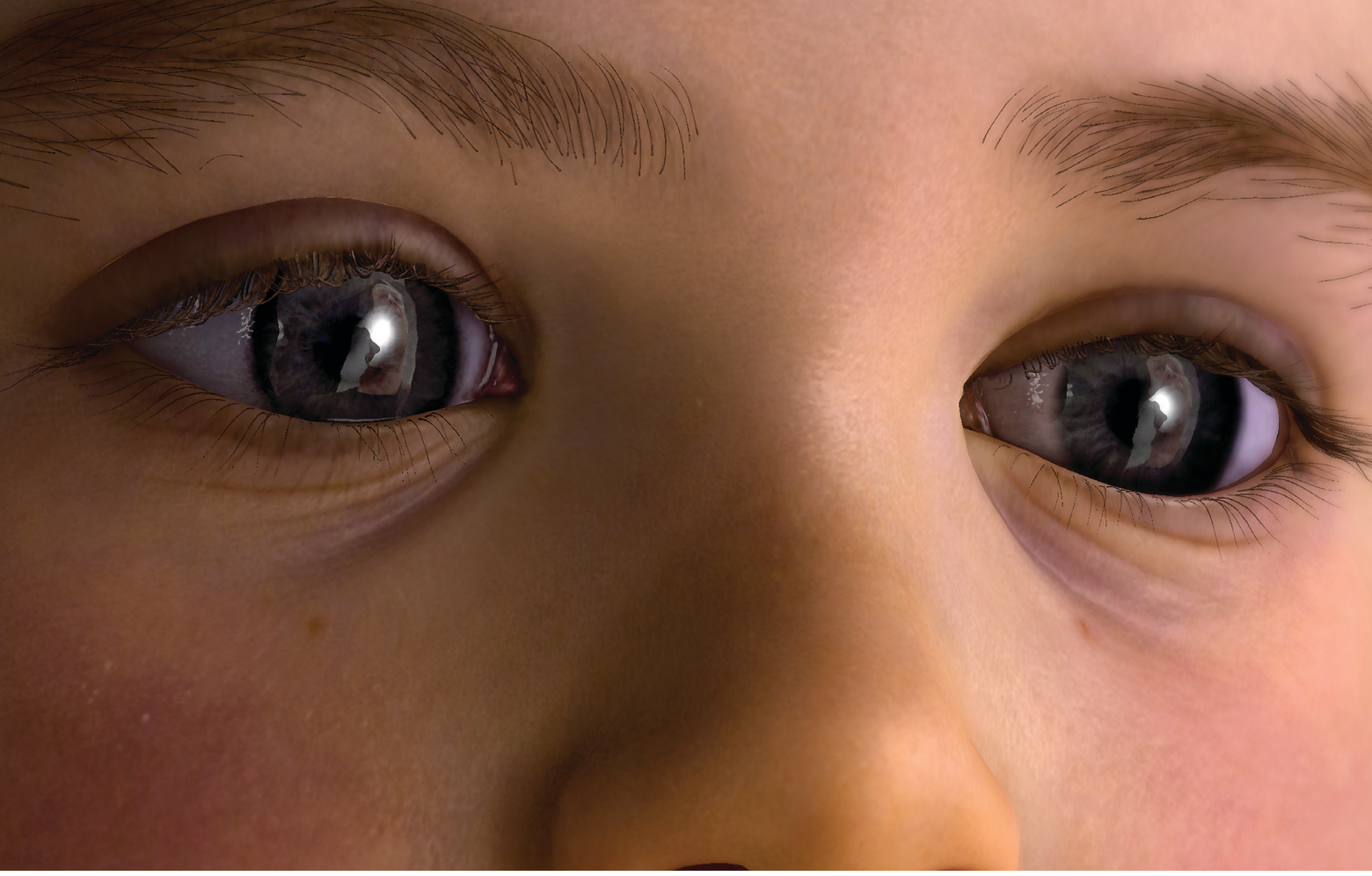
What are the underlying processes from which this most human form of interaction emerges? Will we be able to interact with computers in a face-to-face way that feels natural? This article discusses the unique challenges of realistically simulating the appearance and behavior of the face to create interactive autonomous virtual human models that support naturalistic learning and have the "illusion of life." We describe our recent progress toward this goal with "BabyX," an autonomously animated psychobiological model of a virtual infant. While we explore drivers of facial behavior, we also expect this foundational approach has the potential for more "human" computer interfaces. We also describe our work on our "Auckland Face Simulator" we are developing to broaden this work beyond infants and give a more realistic face and a greater biological basis to adult conversational agents.

Simulating the face has great potential for human-computer interaction (HCI), as it increases the available communication channels between humans and machines in an intuitive, accessible way. But it is also a vehicle with which to explore our own nature. Akin to developmental robotics,[6] which explores ways of learning and mental development through child-like robots, simulating the underlying processes driving the face during social interaction will enable HCI researchers to explore behavioral and learning models involving naturalistic face-to-face interaction.

There is a trend in the game and visual-effects industries to create ever more realistic animated characters, especially humans, but it turns out to not be a straightforward transition from the stylized faces of traditional animation. For these industries, realism of appearance and movement is very important, evidenced by the large financial investment going toward creating the most realistic illusion they can achieve. This is done presumably because the experience becomes more immersive and powerful the closer it is

» key insights

- **The expressive and communicative power of the face has been untapped in HCI but can indeed create deeper human-machine connection and engagement.**
- **The holistic interplay of biologically based computational behavioral models driving a virtual character can give rise to an emotionally affecting experience.**
- **We are developing such a psychobiological modeling framework for autonomous characters and related HCI with realistic yet virtual faces.**



Screenshot of BabyX version 4 (under development) looking at a user; image rendered in real time.

IMAGE COURTESY OF LABORATORY FOR ANIMATE TECHNOLOGIES, THE UNIVERSITY OF AUCKLAND

to reality. Realism lessens the leap required in the suspension of disbelief; it is closer to personal experience. Studies comparing children viewing realistic vs. cartoon depictions of aggression show realism to have an increased affective effect on subsequent behavior.[35]

As characters move to being more human-like, discrepancies in appearance and behavior tend to alarm us, with an increasing chance of falling into the "uncanny valley."[22] The most critical component to "humanness" is the face. Achieving the illusion of life in the face of a realistic digital character is challenging in a passive medium (such as film) but even more so in interactive games and simulations. Concerning realism in computer-generated faces, Gopnik[12] noted: "It made sense to think that the ability to reason and speak was at the heart of the human mind. Turing's bet was that a computer that could carry on a conversation would be convincingly human. But the real imitation game of digital-effects movies suggests that the ability to communicate your emotions may be even more important. The ineffable, subtle, unconscious movements that tell others what we think and feel are what matter most. At least that's what matters most to other human beings."

In an interactive scenario involving unscripted life-like interactions, the problem is even more challenging. The issue is not only how to represent the complex appearance and movement of the synthetic face in real time, but, for an embodied agent's behavior to be believable, it must be consistent and contextually appropriate.[1] If the simulated human can be affected by the interaction, and respond in a way that can affect the user and vice versa, then each partner is more invested in how the interaction unfolds, creating engagement and emotive connection.

What creates all these fleeting movements that communicate so much? And how can a simulation keep them consistent, appropriate, and adaptive? Everything that happens on the face reflects a brain-body state. Because the behavior of the face is affected by so many factors—cognitive, emotional, and physiological—we explore a more-detailed and lower-level biologically based approach than has previously been attempted in facial animation.

Here, we introduce our general approach and design of a modeling framework we call "Brain Language" (BL) to create autonomous expressive embodied models of behavior driven by neural-system models based on affective and cognitive neuroscience theories. Our goal is to integrate different current theories and models to create a holistic "large functioning sketch" of basic aspects of human behavior, with a focus on the face and interactive learning.

Autonomous Animation

There is long-term interest in creating self-animating agents that project the illusion of life. In 1994, Bates[2] described the importance of appropriately timed and clearly expressed emotion to make a character seem alive. Terzopolous[36] introduced a holistic simulation of fish behavior in which each such behavior was driven by an abstracted brain. Terzopolous's work was a closed system, based on initial environmental state, whereas in Maes's ALIVE system,[20] the user was included in the loop with autonomously animated animals, including a virtual

dog "Silas" developed by Blumberg[3] using sophisticated ethological models to simulate how animals are able to organize and adapt.

These and similarly inspired works are important on many levels, as they explain how behavior can emerge, made observable through animation with constraints. Blumberg[3] suggested for a creature to appear alive it must react, have goals, make choices, convey its intentionality, emotionally respond to events, adapt, and vary its movement and response.

For autonomous animation of the face in real time, Terzopolous and Lee[37] developed a physics-based face model driven by a basic behavioral animation model. Despite this pioneering work, few other virtual human studies have focused on this lower level of detail in real-time facial animation.

Most research in building autonomous human agents has been as "embodied conversational agents" (such as in Allbeck[1] and in Cassell[7]) at a generally more phenomenological and higher level, not specifically focused on the subtler details of facial expression and nonverbal behavior; Vinciarelli et al.[39] included a survey of social-signal processing in computer interaction. Simulating these signals is getting greater attention from the interdisciplinary "intelligent virtual agent" community,[5] exploring agents that are capable of real-time perception, cognition, and actions in the social environment; Marsella and Gratch[21] discussed simulations of psychological theories of emotion. Emotion-oriented APIs (such as SEMAINE) have been developed.[33] And Scherer[32] showed how cumulative effects of sequential checks of an eliciting event, mediated by autonomic and somatic components, might combine to create compound facial expressions.

Much of the work on virtual humans has an unfortunately robotic "feeling," particularly with facial interaction. This is possibly due to most virtual human models not focusing on the microdynamics of expression or on facial realism. These microdynamics are considered particularly critical in learning contexts. Rohlfling and Deak[27] stated: "When infants learn in a social environment, they do not simply pick up information passively. They respond to, and learn from, the interaction as they jointly determine its content and quality through real-time contingent and reciprocal coaction."

The dynamic behavior of the face emerges from many systems interacting on multiple levels, from high-level social interaction to low-level biology.

Another important factor in the acceptance of a virtual human face is its visual quality. However, realistically simulating a face—even when still—has proved to be formidable.

Challenges of Modeling the Human Face

The way a digital face moves and appears can cause unwanted effects. Rather than aiding the appearance of life, a partially realistic solution can elicit a negative response—the uncanny valley effect. This response is thought to be triggered by any number of non-expected responses, alarming the viewer's perceptual system.[19] To avoid this response, many factors must be taken into account, including the physical appearance of the face and the eye-gaze movement of the skin, any of which can trigger a form of dissonance that interferes with the affinity of the perceived face.

Appearance. The ability to "read" faces is so important that several different parts of the brain play a role in face perception. We are sensitive to many factors that act as signs of health and vitality. People often refer to someone being "as white as a sheet," "red faced," or "sickly looking"; it is thus important to render physically plausible healthy skin with correct surface properties, detail, and subsurface scattering of light that provides diffuse properties of skin.

This challenge has been approached in two broadly different ways. First, by using "image-based methods" that sample the face under different lighting and viewing conditions[9] and then render the face through a combination of weighted image-blended sets, photogrammetry, and/or image projection. Second, by using "parametric methods" that fit the captured data to a face and material model used during rendering, allowing for more flexibility but at the cost of potentially increased rendering complexity.[16] Given the constraints producing imagery fast enough for user interaction, adding further to the complexity of achieving an effective interactive face, a simplified implementation of the second approach is

typically used for real-time rendering, as in Jimenez.[15]

Deformation. Achieving coherent movement of the skin is especially challenging due to the complex deformations in broad expressions and the highly non-linear motion of skin. Many computer-generated faces in games and films do not address these characteristics; for example, the lips on a character may move while the surrounding areas of the face remain static, causing an unnatural effect. Unlike skeletal muscles, facial muscles are embedded in the mobile facial tissue, meaning facial muscle activation must be treated as a system. Arguably the most coherent and generally useful way to drive facial animation is through parameterization of individual muscle activity (such as in Ekman and Friesen's Facial Action Coding System[10]).

The facial deformations used in animatable faces are typically represented through deforming geometry using weighted joints or weighted shape combination ("blendshape") methods.[25] While effective, these methods can suffer from combinatorial explosion in representing the complex range of facial expressions. The highest-quality models used in the visual-effects industry incorporate a large number of blendshapes to form linear approximations to non-linear deformations. Creating these models is labor intensive, so a number of researchers have approached the problem using flesh simulations.[34,37,40]

Facial Motor System

To design an autonomous digital facial system, it is important to understand how faces are controlled. Traveling inward from the facial nerves, we reach the facial nucleus in the brainstem. The facial nucleus receives its main inputs from both subcortical and cortical areas through different pathways. Both a person's emotional and voluntary facial expressions seem to arise from different neural circuits.[8,13] The implication is that the voluntary expression cannot access a genuine emotional motor pattern and is why it is not possible to fully produce a genuine emotional expression through volition. Similarly, stroke patients with damage to certain primary motor and pre-motor areas cannot produce a symmetrical voluntary smile yet can smile normally in response to jokes.[13]

Expressions are generated by neural patterns in both the subcortical and cortical regions. In the subcortical area, circuits include those for laughing and crying. Evidence suggests certain basic emotional expressions like these do not have to be learned. In comparison, voluntary facial movements (such as those involved in speech and culture-specific expressions) are learned through experience and predominantly rely on cortical motor control.

Our psychobiological facial framework aims to reflect that facial expressions consist of both innate and learned elements and are driven by quite independent brain-region simulations.

Building a Holistic Model

The human face mirrors both the brain and the body, revealing mental state (such as through mental attention in eye direction) and physiological state (such as through position of eyelids and color of the skin). The dynamic behavior of the face emerges from many systems interacting on multiple levels, from high-level social interaction to low-level biology.

To drive a biologically based lifelike autonomous character, one would need to model multiple aspects of a nervous system. Depending on the level of implementation, a non-exhaustive list includes models of the sensory and motor systems, reflexes, perception, emotion and modulatory systems, attention, learning and memory, rewards, decision making, and goals. We seek to define an architecture that is able to interconnect all of these models as a virtual nervous system.

Several biologically inspired cognitive architectures have been developed; see Goertzel et al.[11] for a survey. Most are non-graphical, focusing on cognition over affect or physiological states. It makes sense that the more biologically based the architecture and the more realistic, the more it is ultimately likely to represent biological behavior. An example is the "Leabra" framework,[23] which constructs low-level biologically based neural network models and connects them to model higher-level aspects of cognition. This modeling approach is appealing for its ability to suggest how low-level biological explanations link to high-level behavior (such as goal setting).

Building embodied nervous systems that can learn through real-time sensorimotor interaction is being explored in the field of developmental robotics.[8] Social-interaction models have been explored with anthropomorphized "social robots" (such as in Leonardo and Kismet[4]). Developmental robotics, in particular, seeks to explore the theory of embodied cognition—how the mind develops through real-time sensorimotor interaction.

Our approach to building live interactive virtual agents takes a similar direction whereby we embody, through realistic computer graphics, a biologically based model of behavior. We ground experience through interaction and place particular emphasis on the importance of face-to-face interaction, which is difficult to achieve in robotics due to mechanical constraints. The result is a system that can be reduced to more biological detail, as well as expandable to incorporate higher-level complex systems. As there are many competing theories on how different brain and behavioral systems function, our choice is to opt for flexibility and develop a "system to build systems" in a Lego-like manner.

Brain Language

BL[28] is a modular simulation framework we have been developing for the past five years to integrate neural networks with real-time computer graphics and sensing. It is designed for maximum flexibility and can be connected with other architectures through a simple API. It consists of a library of time-stepping modules and connectors. It is designed to support a wide range of computational neuroscience models, as in Trappenberg.[38] Models supported by BL range from simple leaky integrators to spiking neurons to mean field models to self-organizing maps. These can be interconnected to form larger neural networks (such as recurrent networks and convolutional networks like those used in deep learning). Our main interest is in online learning, in which the network learns during live interaction from both spatial and temporal data. A key strength of BL is its tight integration with computer graphics as a visualization tool. Complex dynamic

Figure 1. BabyX's interactive brain: (left) superior colliculus activity driving visual attention is visible (green) in the brainstem; (middle) BL raster plot of neural activity and scrolling display of modulatory activity; (right) basal ganglia circuit and interactive dopamine level modification (green) affecting cortico-thalamic feedback and eye movement.

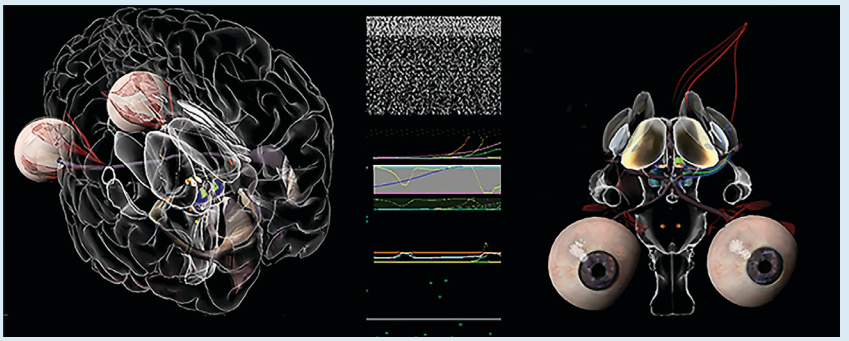
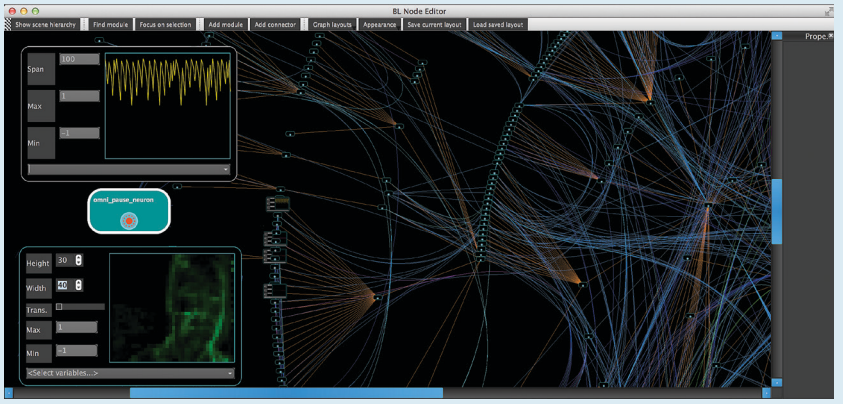


Figure 2. Screenshot of interactive BL viewer: (left) BL scopes viewing activity of a single neuron (top) or an array of retinotopic neurons (bottom) during a live interaction; (right) partial view of BabyX's virtual connectome, which can be explored interactively; connections light up (green or red) when activated.



networks can be visually investigated in multiple ways, either through 3D computer graphics (see Figure 1) or through a 2D schematic interface showing activity on the virtual connectome (see Figure 2). Individual variable activity (such as neuron voltage or firing rate) can be inspected with scopes during a simulation, as in Figure 2 left. The simulation can be interactively modified (such as by changing neural-model parameters while viewing the effect on the animation) (see Figure 3 right).

Sensory input is typically through camera, microphone, and keyboard to enable computer vision audition and "touch" processing, but data can be input from any arbitrary sensor or output to an effector through the API. Computer graphics output is through OpenGL and the OpenGL Shading Language. A key feature of BL is that any variable in the neural network system can be shared and drive any aspect of a sophisticated 3D animation system; for more detail on BL, see Sagar et al.[28]

BabyX Project

To illustrate how these concepts come together, we describe an experimental psychobiological simulation of an infant we call "BabyX,"[29,31] that aims to embody models of interactive behavior and social learning to create an autonomous virtual infant one can interact with naturally.

Facial expression. At a conceptual level, BabyX's computer-graphic face model is driven by muscle activations generated from motor-neuron activity. The facial expressions are created by modeling the effect of individual muscle activations and their non-linear combination forming her range of expressions, as in Figure 3. The modeling procedures involve biomechanical simulation, scanning, and geometric modeling. Fine details of visually important elements (such as the mouth, eyes, eyelashes, and eyelid geometry) are painstakingly constructed for life-like reality (see Figure 4).

A highly detailed biomechanical face model, as in Figure 3, has been constructed from MRI scans and anatomic reference, akin to Wu.[40] Skin deformation is generated by individual or grouped-muscle activations. We have modeled the deep and superficial fat, as well as muscle, fascia, connective tissue, and their various properties. We have used large-deformation finite-element elasticity[40] to deform the face from rest position through simulated muscle activation. Individual and combined muscle activations were simulated to form an expression space,[30] interpolated on the fly in BL as the face animates. The response to muscle activation is consistent skin deformation and motion.

Nervous system. BabyX's biologically inspired nervous system consists of an interconnected set of neural system and subsystem models. The models implemented so far are sparse yet span the neuroaxis and generate muscle-activation-based animation as motor output from continuously integrated neural network models. Due to the Lego-like nature of BL, we can have a closed-loop functioning system allowing experimental interchange of components while exploring different theoretical models. In total, the models aim to form a "large functioning sketch" of interconnected mechanistic systems contributing to behavior, containing both top-down and bottom up mechanisms interacting as an integrated system.

BabyX's neural networks and circuits implemented so far cover basic elements of motor control, behavior selection, reflex actions, visual attention, learning, salience, emotion, and motivation. An architectural diagram relating some of the key functional components, neuroanatomical structures, and functional loops is included in Figure 5; note cortical and subcortical input to the facial nucleus. A characteristic of this modeling approach is the representation of subcortical structures (such as the basal ganglia) and brainstem nuclei (such as the oculomotor nuclei). The structures are functionally implemented as neural

network models with particular characteristics (such as the amygdala and hippocampus as "hebbian associators" and the pulvinar as a topographically organized array of neurons forming a saliency map). Activation of the hypothalamus releases virtual hormones. Cortical regions use recurrent and multi-layer neural networks and self-organizing maps. Due to the Lego-like nature of BL, simple models can be replaced by more sophisticated models as they become available.

One of the goals of BabyX is to visually represent functional neural-circuit models in their appropriate anatomical positions. For example, our Basal Ganglia model (based on Redgrave et al.[26]) controls motor actions and has an appropriate 3D location and geometry, as in Figure 1, and the activity of the specific neurons form inputs to the shaders to show the circuit in action as it processes.

Emotions in BabyX are, in fact, coordinated brain-body states that modulate activity in other circuits (such as increasing the gain on perceptual circuits). Emotional states modulate the sensitivity of behavioral circuits. For example, stress lowers the threshold for triggering a brainstem central pattern generator that, in turn, generates the motor pattern of facial muscles in crying.

Neurotransmitters and neuromodulators play many key roles in BabyX's learning and affective systems.[24] An example of a physiological variable that affects both the internal and external state of BabyX is dopamine, which provides a good example of how modeling at a low level interlinks various phenomena. In BabyX, virtual dopamine plays a key role in motor activity and reinforcement learning. It can also modulate plasticity in the neural networks and have subtle behavioral effects such as pupil dilation and blink rate. The use of such low-level models means the user can adjust BabyX's behavioral dynamics, sensitivities, and even temperament by adjusting virtual neurotransmitter levels.

Sensory input. BabyX takes audio-visual input Web camera and microphone, and "touch" from keyboard or touchscreen and is designed to work without special hardware. BL can interface to different devices, and the BabyX project exists separately from choice of display systems (such as virtual reality, or VR, or augmented reality, or AR). Advances in AR, particularly in systems that allow for facial-expression tracking, accurate eye tracking, and depth gaze registration of the user, mean an obvious possible implementation could be as a virtual agent in AR. Such an additional level of engagement would enhance the experience but also benefit from the tight emotional signaling feedback we have developed.

Learning through interaction. One motivation for the BabyX project is to ex-

Figure 3. BabyX (version 1). Detailed biomechanical face model simulating expressions generated from muscle activations.

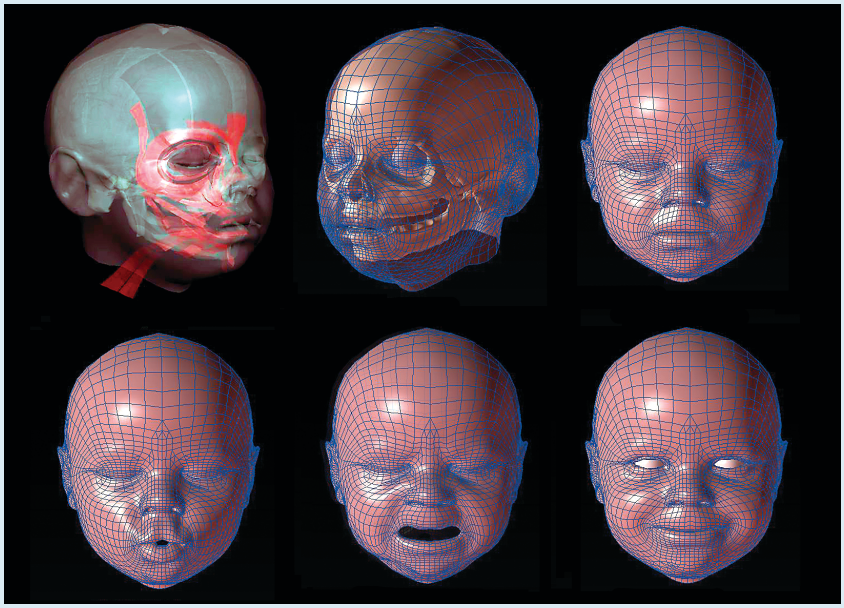


Figure 4. BabyX (version 4, under development). Screenshot from real-time interactive psychobiological virtual infant simulation.

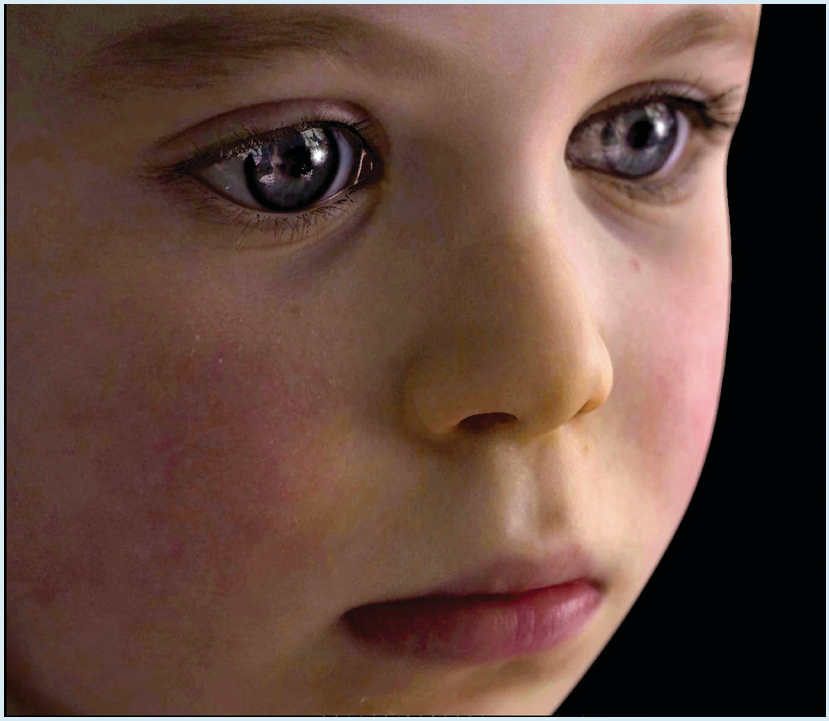
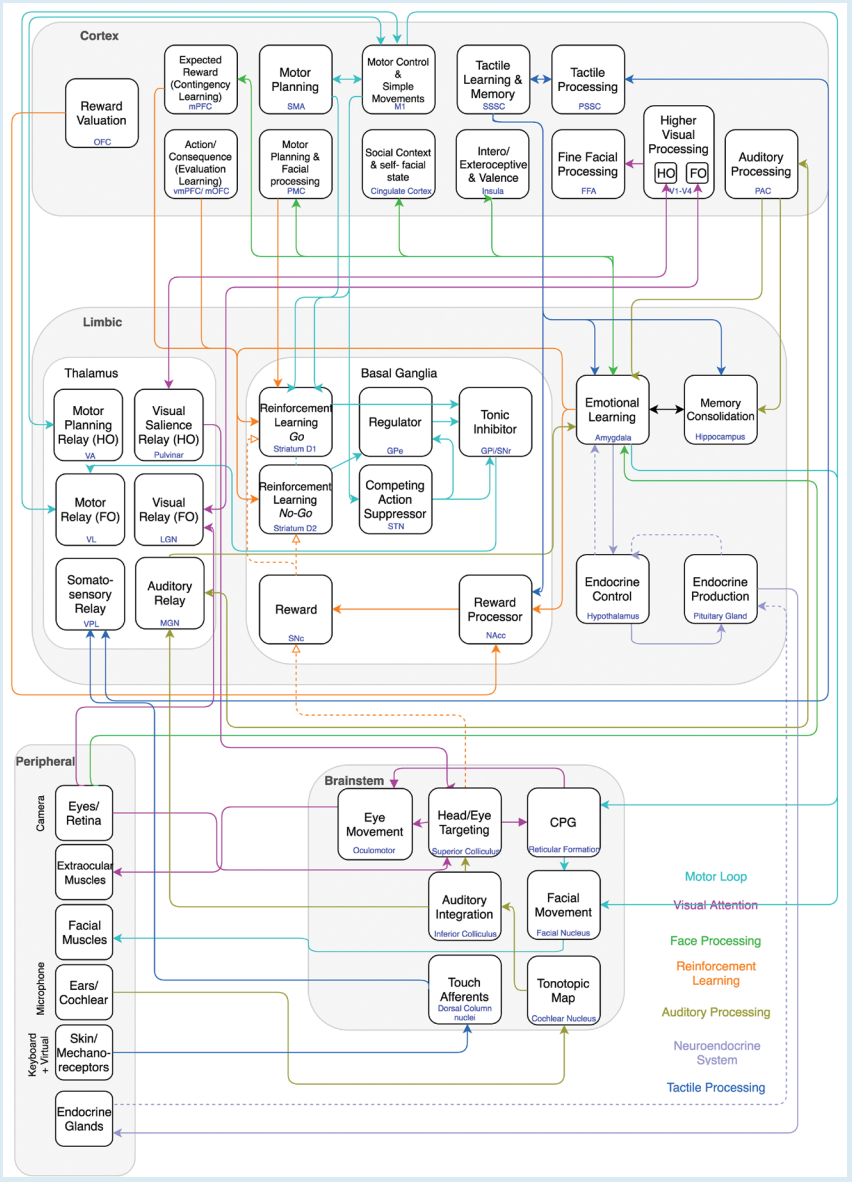


Figure 5. Architectural diagram showing several key functional components and processing loops and their neuroanatomical equivalents (blue text).



plore how high-level social interactions might interact with models of lower-level biological mechanisms. An example illustrating such interaction we have investigated is facial mimicry (see Figure 6); for a related video, see https://vimeo.com/123986611. This may be simple to program at a high level in an embodied agent (such as by copying inputs to outputs), but when exploring how the interaction may emerge in a general sense, through biologically plausible intrinsic learning mechanisms, facial mimicry raises complex questions. It is a key example cited in the mirror-neuron debate.[14] How do we learn to map other people's actions to our own? The mirror system is fundamental to learning, but how could the mapping between an infant's expressions and those of a caregiver occur?

One possible mechanism is through "associative sequencing"[14] in which spontaneous motor activity causes a facial action that becomes associated with sensory input caused by the caregiver's response. Spontaneous "motor babbling" activity is thought to be a fundamental way to bootstrap exploration of motor space and considered fundamental to the development of autonomous agency.[18] In our model, babbling is generated by spontaneous neural network activity, modulated by physiological parameters leading to activation of facial motor pattern generators. BabyX may spontaneously make a puckered mouth shape, which is seen and then mirrored by the user or "caregiver." For BabyX, the activity and feedback from this facial movement is associated with a delayed visual sensory pattern from the caregiver's facial response. Associative learning is modulated by phasic dopamine. Strengthening bidirectional associative connections results in the caregiver's expression being able elicit a similar expression in BabyX. On successful imitation of an expression, the caregiver may praise with positive affect, releasing yet more dopamine, which further strengthens synaptic connections. The sensed positive affect and dopamine release activate and modulate affective circuits, causing neural-pattern generators to contract the muscles for smiling. For BabyX to learn an association of her muscle activity to the caregiver's expressions, she must attend to the caregiver's face. The caregiver's actions cause activity on BabyX's superior and inferior colliculus, where sensory events compete to drive the oculomotor network to move the eyes. A subregion of the camera input is automatically mapped to the virtual "fovea," which maps to where the eyes are focusing.

In sum, the various circuits driving BabyX's facial expressions converge on the facial nucleus in the brainstem that then activates BabyX's animated facial muscles. Because the various inputs to the facial nucleus arise from the activity of independent yet interconnected networks, BabyX's facial expressions can be understood in the context of internal activity and external factors.

While BabyX's "default" expressions are associated with reflexes and basic affective states (such as newborns crying), the learned expressions here are voluntary. Dating to Charles Darwin, the "facial feedback hypothesis" states posing an expression can influence emotional state. By adding bidirectional connections to the affective networks that generated them, mimicked expressions can cause activity in BabyX's affective circuits, functioning perhaps as a basis for "virtual empathy."

Action discovery. An example showing learning through interaction with the environment in order to demonstrate autonomous action discovery[26] was to aim to have BabyX learn to play the classic video game "Pong" (see Figure 7). We thus connected motor neurons in BabyX to the bat controls and overlaid the visual output of the game on the camera's input. Motor babbling causes the virtual infant to inadvertently move the bat, much like a baby might flail its arms about. Trajectories of the ball are learned as spatiotemporal patterns on neural network maps. If the bat hits the ball, a rewarding reaction results, reinforcing the association of the current motor state with the trajectory. This association further results in the bat being moved in anticipation of where the ball is going. Without further modification to the model, it is possible for the user to actively encourage BabyX's choices (releasing virtual dopamine), providing a nice example of "naturally supervised" reinforcement learning.

These basic examples show BabyX learning through interaction with a human user and the shared environment. While basic, these examples of intrinsic action discovery, association, and reinforcement learning (unsupervised and "naturally" supervised) are fundamental to developing generalized autonomous learning systems.

Observations. As interaction is central to the phenomena, we have demonstrated and tested BabyX in several public forums where we have observed an extension of emotion from BabyX to a shared experience with a "passive" audience reacting as vicarious participant. Audience behavior is absorbed into the simulation and is not apart from it, making the experience different from a film, game, or pre-rendered simulation. The demonstrator elicits behavior from BabyX through visual and vocal activity and tries to direct her attention. Affective expressions and voice stimulate reward and affective circuits. If BabyX is abandoned, depending on oxytocin levels, her stress system can activate a cascade of virtual hormones, and she becomes increasingly distressed. BabyX can be trained to recognize certain images that can be associated with vocalizations. When the demonstrator gains BabyX's attention and shows her a "First Words Book," if an image causes a strong enough activation, the image will trigger BabyX to voice an associated word.

Observing in a real, unscripted environment, people anticipate and seek emotional responses from BabyX. As such engagement happens, they are often transformed from observers to engaged participants. An example of this was seen at the 2015 SIGGRAPH conference where Sagar[31] demonstrated BabyX. While the audience was mainly informed professionals, their reaction was audible and visceral to BabyX. Responses repeatedly observed included a sharp negative reaction when the demonstrator offered to demonstrate the pain response, as if someone was about to "hurt" the baby. There was no sense that the audience rationally thought the baby was real, though it immediately reacted as if she were. Even within a formal academic presentation, and with the pain response a valid part of any brain model, the audience reacted as if the demonstrator was about to be cruel. Interestingly, this was followed by an emotional display of relief in the form of laughter. As laughter is infectious, the demonstrator laughed, which was registered by BabyX's sensory inputs, causing her to be "happier." The audience thus became a part of the feedback loop that changed both parties' emotional states. The implication is that a witness to a BabyX session is to

Figure 6. BabyX (version 3). Screenshot of sensorimotor online learning session in which multiple inputs and outputs of the model can be viewed simultaneously, including scrolling displays, spike rasters, plasticity, activity of specific neurons, camera input, animated output.

Figure 7. BabyX (version 3). Learning to play the video game "Pong" through action discovery and online reinforcement learning.

become a part of the holistic environment and the interactive experience. There may be ethical implications as well, and further research is needed to investigate the co-defined dynamic interaction that allows such strong "in the moment" emotional responses, as such responses may have long-term interface implications.

Further development and validation. We are currently working on BabyX version 4, as in Figure 4, which has a virtual body and is able to control her limbs, with initial focus on learning to reach and grasp. BabyX version 4 is intended to interact with the public in exhibitions, performing basic learning tasks (such as label learning). For speech, BabyX babbles with a synthesized voice sampled from phonemes produced by a real child. We are implementing techniques so BabyX can learn an acoustic mapping from any arbitrary voice to construct new words using her own voice. Lip shapes are pre-associated with acoustic elements. Our aim for BabyX is that she should be capable of learning arbitrary sensorimotor sequences, theorized to map to sentence construction.[17]

In an ongoing developmental psychology study, we are conducting a detailed quantitative characterization of the microdynamics of early social learning between parents and their infants. As a first step to validate the effectiveness of BabyX's behavior at a high level, we will be exploring how well the model elicits naturalistic responses from parents in a social interaction loop, compared to their own or another child. If the model is successful, we will have a new way to study coordinated interaction, and how the way in which we teach infants may play a critical role in learning. Introducing synthetic lesions could be an effective way to explore lower-level validation.

Figure 8. The Auckland Face Simulator is being developed to create realistic and precisely controllable real-time models of the human face and its expressive dynamics for psychology research and real-time HCI applications.

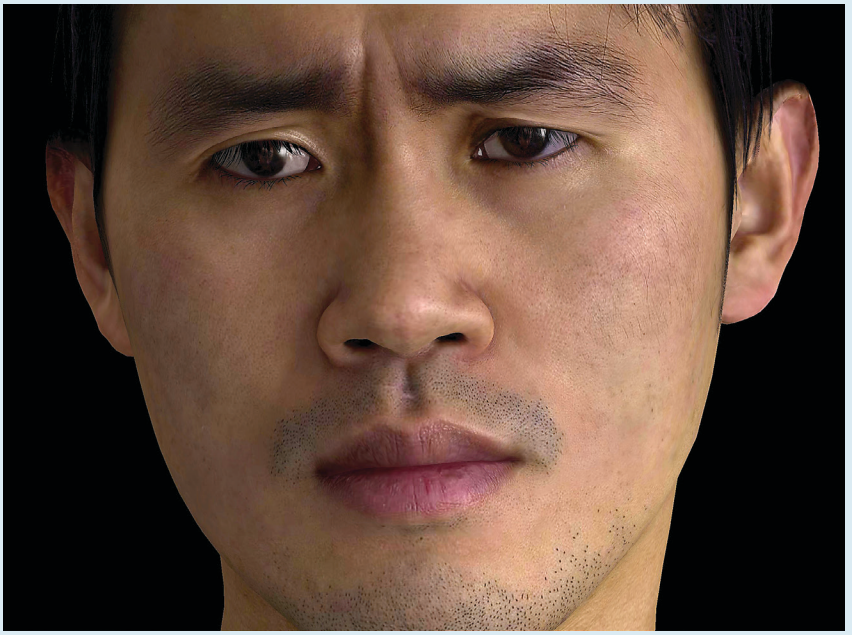


Figure 9. The Auckland Face Simulator enables autonomously animated faces to be used for cinematic-like extreme close-up shots.

The Auckland Face Simulator

A developing infant is certainly not the easiest approach for creating an embodied conversational agent for HCI tasks. For this purpose, we are building on the same underlying computational platform the "Auckland Face Simulator" (see Figure 8 and Figure 9, as well as Figure 4) also demonstrated at SIGGRAPH[31] to produce highly realistic avatars capable of real-time interaction; for a related video, see https://vimeo.com/128835008. These faces are designed to be used as stimuli for psychological research but also to provide a realistic interface for third-party virtual-agent and AI applications. The avatars can be "told what to say" using text to speech (TTS), and the nonverbal behavior can be specified in a simple API or custom TTS markup language to add further meaning. Wrinkling the nose or raising the upper lip while speaking can dramatically change the perceived meaning. BL allows internal variables of the avatar's nervous system to be controlled at any level, from muscles to affective circuits.

Conclusion

Engaging face to face with an interactive computer model requires autonomy with contextual responsiveness. If visually consistent, realistic appearance and movement seem to increase the sensory intensity of the experience. Internally consistent generative models enable cognitive, affective, and physiological

factors that drive facial behavior to be produced coherently, justifying a lower-level more biologically based modeling approach than has previously been taken with virtual human faces. Exploring these elements together allows new yet familiar phenomena to occur. New, because we do not normally experience this sort of interaction with computers, familiar because we do with people.

Being able to simulate the underlying drivers of behavior, realistic appearance and real-time interaction together deliver three aspects of interaction, but virtually:

Explore. Allows us to explore how the interplay of biologically based systems can give rise to an emotionally affecting experience on a visceral, intuitively relatable human level;

Include movements. Applies an embodied-cognition approach to include the subtle and unconscious movements of the face as a crucial part of mental development and social learning; and

Understand key requirements. Gives a basis for understanding the key requirements for more natural and adaptive HCI in which the interface has a face.

The virtual infant BabyX is not an end unto itself but allows researchers to study and learn about the nature of human response. There is a co-defined dynamic interaction where one can adjust to BabyX no longer as a simulation but as a personal encounter.

In summary, the enormous complexity of modeling human behavior and dyadic interaction cannot be overestimated, but naturalistic autonomous virtual humans who embody and process theoretical models of our behavior and reflect them back at us may give us new insight into core aspects of our nature and interaction with other people—and future machines.

Acknowledgments

This work was supported in part by the University of Auckland Vice-Chancellor's Strategic Development Fund, Cross Faculty Research Initiative Fund, Strategic Research Investment Fund, and Ministry of Business Innovation and Employment "Smart Ideas" program. We also thank Kieran Brennan, Stephanie Khuu, Kai Riemer, and John Reynolds.

References

1. Allbeck, J. and Badler, N. Consistant communiction with control. In *Proceedings of the Workshop on Multimodal Communication and Context in Embodied Agents at the Autonomous Agents Conference* (2001).
2. Bates, J. The role of emotion in believable agents. *Commun. ACM 37*, 7 (July 1994), 122–125.
3. Blumberg, B.M. *Old Tricks, New Dogs: Ethology and Interactive Creatures.* Ph.D. thesis, MIT, Cambridge, MA, 1996.
4. Breazeal, C. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies 59*, 1 (2003), 119–155.
5. Brinkman, W.P., Broekens, J., and Heylen, D., Eds. *Proceedings of the Intelligent Virtual Agents: 15th International Conference* (Delft, The Netherlands, Aug. 26–28). Springer, 2015.
6. Cangelosi, A., Schlesinger, M., and Smith, L.B. *Developmental Robotics: From Babies to Robots.* MIT Press, Cambridge, MA, 2015.
7. Cassell, J. *Embodied Conversational Agents.* MIT Press, Cambridge, MA, 2015.
8. Cattaneo, L. and Pavesi, G. The facial motor system. *Neuroscience & Biobehavioral Reviews 38* (2014), 135–159.
9. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., and Sagar, M. Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (New Orleans, LA, July 23–28). ACM Press/Addison-Wesley Publishing Co., New York, 2000, 145–156.
10. Ekman, P. and Friesen, W.V. *Facial Action Coding System: Investigator's Guide Part I.* Consulting Psychologist Press, Palo Alto, CA, 1978.
11. Goertzel, B., Lian, R., Arel, I., De Garis, H., and Chen, S. A world survey of artificial brain projects. Part II: Biologically inspired cognitive architectures. *Neurocomputing 74*, 1 (2010), 30–49.
12. Gopnick, A. Why digital-movie effects still can't do a human face. *The Wall Street Journal* (Jan. 8, 2015).
13. Gothard, K. The amygdalo-motor pathways and the control of facial expressions. *Frontiers In Neuroscience 8* (2014).
14. Heyes, C. Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews 34*, 4 (2010), 575–583.
15. Jimenez, J., Sunstedt, V., and Gutierrez, D. Screen-space perceptual rendering of human skin. *ACM Transactions on Applied Perception 6*, 4 (2009), 23.
16. Klehm, O. et al. Recent advances in facial appearance capture. *Computer Graphics Forum 34*, 2 (2015), 709–733.
17. Knott, A. *Sensorimotor Cognition and Natural Language Syntax.* MIT Press, Cambridge, MA, 2012.
18. Lee, M.H. Intrinsic activity: From motor babbling to play. In *Proceedings of the IEEE International Conference on Development and Learning* (Frankfurt am Main, Germany, Aug. 24–27). IEEE Press, 2011.
19. MacDorman, K. and Entezari, S. Individual differences predict sensitivity to the uncanny valley. *Interaction Studies 16*, 2 (2015), 141–172.
20. Maes, P. Artificial life meets entertainment: Lifelike autonomous agents. *Commun. ACM 38*, 11 (Nov. 1995), 108–114.
21. Marsella, S. and Gratch, J. Computationally modeling human emotion. *Commun. ACM 57*, 12 (Dec. 2014), 56–67.
22. Mori, M., MacDorman, K.F., and Norri, K. 'The uncanny valley.' *Robotics & Automation Magazine 19*, 2 (2012), 98–100.
23. O'Reilly, R., Hazy, T., and Herd, S. The leabra cognitive architecture: How to play 20 principles with nature and win! In *Oxford Handbook of Cognitive Science*, S. Chipman, Ed., Oxford University Press, Oxford, U.K., 2012.
24. Panksepp, J. *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford University Press, 1998.
25. Parke, F. and Waters, K. *Computer Facial Animation.* CRC Press, 2008.
26. Redgrave, P., Gurney, K., and Reynolds, J. What is reinforced by phasic dopamine signals? *Brain Research Reviews 58*, 2 (2008), 322–339.
27. Rohlfling, K. and Deak, G. Microdynamics of interaction: Capturing and modeling infants' social learning. *IEEE Transactions on Autonomous Mental Development 5*, 3 (Sept. 2013), 189–191.
28. Sagar, M., Robertson, P., Bullivant, D., Efimov, O., Jawed, K., Kalarot, R., and Wu, T. BL: A visual computing framework for interactive neural system models of embodied cognition and face-to-face social learning. In *Proceedings of the 14th International Conference on Unconventional Computation and Natural Computation* (Auckland, New Zealand). Springer, Heidelberg, Germany, 2015, 71–88.
29. Sagar, M., Bullivant, D., Robertson, P., Efimov, O., Jawed, K., Kalarot, R., and Wu, T. A neuro-behavioural framework for autonomous animation of virtual human faces. In *Proceedings of SIGGRAPH Asia Autonomous Virtual Humans and Social Robots for Telepresence* (Shenzhen, China, Dec. 3–6). ACM Press, New York, 2014.
30. Sagar, M. Facial performance capture and expressive translation for King Kong. In *Proceedings of ACM SIGGRAPH 2006 Sketches* (Boston, MA, July 30–Aug. 3). ACM, Press, New York, 2006, 26.
31. Sagar, M. BabyX and the Auckland Face Simulator. In *Proceedings of ACM SIGGRAPH Computer Animation Festival* (Los Angeles, CA, Aug. 9–13). ACM Press, New York, 2015, 183–184.
32. Scherer, K, Mortillaro, M., and Mehu, M. Understanding the mechanisms underlying the production of facial expression of emotion: A componential perspective. *Emotion Review 5*, 1 (2013), 47–53.
33. Schröder, M. The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems. *Advances in Human-Computer Interaction* (2010).
34. Sifakis, E., Neverov, I., and Fedkiw, R. Automatic determination of facial muscle activations from sparse motion-capture marker data. *ACM Transactions on Graphics 24*, 3 (July 2005), 417–425.
35. Stone, R. and Hapkiewicz, W. The effect of realistic versus imaginary aggressive models on children's interpersonal play. *Child Development 42*, 5 (1971), 1583–1585.
36. Terzopoulos, D. et al. Artificial fishes with autonomous locomotion, perception, behavior, and learning in a physical world. In *Proceedings of the Artificial Life IV Workshop*, P. Maes and R. Brooks, Eds. (Cambridge, MA, July 6–8). MIT Press, Cambridge, MA, 1994.
37. Terzopoulos, D. and Lee, Y. Behavioral animation of faces: Parallel, distributed, and real-time facial modeling and animation. In *Proceedings of ACM SIGGRAPH* (Los Angeles, CA, Aug. 8–12). ACM Press, New York, 2004, 119–128.
38. Trappenberg, T. *Fundamentals of Computational Neuroscience.* Oxford University Press, New York, 2010.
39. Vinciarelli, A. et al. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing 3*, 1 (2012), 69–87.
40. Wu, T. A *Computational Framework for Modeling the Biomechanics of Human Facial Expressions.* Ph.D. thesis, The University of Auckland, Auckland, New Zealand, 2014.

Mark Sagar (m.sagar@auckland.ac.nz) is an associate professor in the Auckland Bioengineering Institute and director of the Laboratory for Animate Technologies at the University of Auckland, Auckland, New Zealand, and CEO/founder of Soul Machines Ltd., Auckland, New Zealand.

Mike Seymour (mike.seymour@sydney.edu.au) is a lecturer in information systems at the University of Sydney, Sydney, Australia.

Annette Henderson (a.henderson@auckland.ac.nz) is a developmental psychologist and senior lecturer in the School of Psychology at the University of Auckland, Auckland, New Zealand.

BabyX and Auckland Face Simulator research and development contributors:
David Bullivant (d.bullivant@auckland.ac.nz),
Paul Corballis (p.corballis@auckland.ac.nz),
Oleg Efimov (oefi712@auckland.ac.nz),
Khurram Jawed (mjaw002@auckland.ac.nz),
Ratheesh Kalarot (rkal018@auckland.ac.nz),
Paul Robertson (prob014@auckland.ac.nz),
Werner Ollewagen (woll627@auckland.ac.nz), and
Tim Wu (twu051@auckland.ac.nz), all at the University of Auckland, Auckland, New Zealand.

Watch the authors discuss their work in this exclusive *Communications* video. http://cacm.acm.org/videos/creating-connection-with-autonomous-facial-animation