This is the accepted manuscript for: Proudfoot D. (2013) Can a Robot Smile? Wittgenstein on Facial Expression. In Racine TP; Slaney KL (Ed.), *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*: 172-194. New York: Palgrave Macmillan. DOI: 10.1057/9781137384287_10.

In citing this manuscript, please refer to the published version.

Can a Robot Smile? Wittgenstein on Facial Expression

Diane Proudfoot
Department of Philosophy and the Turing Archive for the History of Computing
University of Canterbury, New Zealand

Abstract Some researchers in social robotics aim to build 'face robots'—machines that interact with human beings (or other robots) by means of facial expression and gesture. They aim, in part, to use these robots to test hypotheses concerning human social and psychological development (and disorders such as autism) in controlled, repeatable experiments. A robot may be said to 'grin' and 'frown', or to have 'a smile on its face'. This is not to claim merely that the robot has a certain physical configuration or behaviour; nor is it to say merely that the robot's 'facial' display is, like an emoticon or photograph, a *representation* of a smile or frown. Although researchers may refrain from claiming that their machines have *emotions*, they attribute *expressive behaviours* to them literally and without qualification. Wittgenstein said, however, 'A smiling mouth *smiles* only in a human face'. Smiling is a complex conventional gesture. A facial display is a *smile* only if it has a certain meaning—the meaning that distinguishes a smile from a human grimace or facial tic, and from a chimpanzee's bared-teeth display. In this paper I explore the implications of Wittgenstein's remarks on expression for the claim that face robots can smile or frown.

1. Introduction

Recent work in social robotics, which is aimed both at creating an artificial intelligence and providing a test-bed for psychological theories of human social development, involves building robots that can learn from 'face-to-face' interaction with human beings—as human infants do. The building-blocks of this interaction include the robot's 'expressive' behaviours, for example, facial expression and head-and-neck gesture. There is here is an ideal opportunity to apply Wittgensteinian conceptual analysis to current theoretical and empirical work in the sciences. Wittgenstein's philosophical psychology is sympathetic to embodied and situated Artificial Intelligence (see Proudfoot, 2002, 2004b), and his discussion of facial expression is remarkably modern. In this paper I explore his approach to facial expression, using smiling as a representative example, and apply it to the canonical interactive face robot, Cynthia Breazeal's Kismet (see e.g. Breazeal, 2009, 2002). I assess the claim that Kismet has expressive behaviours, with the aim of generating philosophical insights for AI.

Section 2 describes aspects of recent work in the subfield of social robotics that I call *face AI*. Sections 3 and 4 analyse and elaborate on Wittgenstein's remarks concerning facial expression and other expressive behaviour, in order to answer the question *Can a face robot smile?*. I unpack the significance of Wittgenstein's remark, 'A smiling mouth *smiles* only in a human face' (PI §583). Smiling is more than the production of a certain physical configuration or behaviour. It is a complex conventional gesture. A facial display is a *smile* only if it has a certain meaning—the meaning that distinguishes a smile from a human grimace or facial tic, and from a chimpanzee's bared-teeth display. According to Wittgenstein, a display acquires this meaning only if it is part of a 'normal play' of mobile human expressions and is appropriately embedded in the 'bustle of life'. To be *smiles*, the displays of face robots must satisfy this requirement.

Section 5 applies Wittgenstein's remarks on expressive behaviour to human infants, and compares the infant with the 'child-machine'—a machine that is to learn human-like behaviours autonomously, aided by human scaffolding. Although neither the infant's nor the child-machine's 'smile' suffices for *smiling*, the infant's display is closer to satisfying the conceptual requirements on expressive behaviour. Section 6 analyses Wittgenstein's remarks on the recognition of facial expressions and on (what is now called) mind-blindness. This yields a philosophical characterization of an 'expressive' face robot as a *mind-blind smiling-machine.*

A Wittgensteinian perspective on the claims of ('strong') face AI makes three contributions. First, it generates new conceptual claims about expressive behaviour and facial-expression recognition, which are also the bases of testable empirical hypotheses. Second, it provides a conceptual framework that is valuable in tackling (what I call) the forensic problem of anthropomorphism for AI (Proudfoot, 2011). Our tendency to anthropomorphism stands in the way of AI's grand goal of building a physical machine that thinks; any supposed demonstration of artificial intelligence must exclude the possibility that we are merely anthropomorphizing an unintelligent machine. Wittgenstein's distinction between different senses of 'smiling' ('frowning', and so on) helps to avoid the misplaced anthropomorphism of social robots. Third, Wittgenstein's ideas point to crucial differences between the human child and the *child-machine*, a fundamental concept in machine intelligence since Turing.

2. Face AI

Despite the confidence of some early researchers, 'Good Old-Fashioned AI' failed to produce a human-like artificial intelligence, and ran into seemingly intractable problems—for example, the difficulty of building a system capable of 'common-sense' knowledge. This led in the 1980s and 1990s to a greater emphasis upon embodied and socially situated AI (see e.g. Brooks, 1999). The grand aspiration of building what Turing called a child-machine (Turing, 1950) includes implementing on a robot 'theory of mind' abilities, such as face and agent detection, joint visual attention, self-recognition, and success in false belief tasks (e.g. Scassellati, 2000, 2002; Gold and Scassellati, 2007; Breazeal, Gray, and Berlin, 2009). Other abilities of pre-verbal infants to be implemented include the detection of prosody in a human voice and of intent in visual motion (Scassellati et al., 2006; Kim and Scassellati, 2007; Breazeal and Aryananda, 2002), and arm reaching and imperative pointing (e.g. Sun and Scassellati, 2005; Scassellati, 2000). Social roboticists aim to use results from cognitive neuroscience and developmental psychology to implement these basic behaviours on a robot, and then to bootstrap more complex behaviour using strategies analogous to those of human infant-carer (or sibling-sibling) interaction (see e.g. Breazeal, 2009). The goal is a 'socially intelligent' robot (for an overview of the principles underlying this work in human-robot interaction, see Dautenham, 2007). The robots involved include Cog (e.g. Brooks et al., 1999), Kismet (Breazeal, 2002), Leonardo (Breazeal, 2006, 2009; Breazeal et al., 2005), Nexi (Breazeal, 2009), Bandit (e.g. Wade et al., 2012), Nico (Sun and Scassellati, 2005, Scassellati et al., 2006), Infanoid (Kozima, Nakagawa, and Yano, 2005), and Mertz (Aryananda, 2006).

Researchers in social robotics also have narrower aims. These include: testing hypotheses about human social and psychological behaviour and development (and about cognitive developmental disorders such as autism), by investigating human-robot interaction in standardized conditions (e.g. Scassellati, 2005); building artificial systems that are capable of quickly learning new behaviours from (and responding accurately and sensitively to) humans, by means of normal social cues (e.g. Breazeal, 2006); and producing service, entertainment, and therapeutic robots, with which humans with no specialized training can interact intuitively (Scassellati, 2000, 2005; Tapus, Matarić, and Scassellati, 2007; Fasola and Matarić, 2012; Wade et al., 2012; Feil-Seifer and Matarić, 2011; Giullian et al., 2010). The study of human-robot interaction and of 'expressive' anthropomorphic robots has expanded hugely in recent years (for a taxonomy of social robots and of design methodologies, see Fong, Nourbakshsh, and Dautenham, 2003).

Turing suggested in 1950 that one approach to machine intelligence would be to provide a machine with 'the best sense organs that money can buy', and then 'teach it to understand and speak English' (1950, p. 460). Modern roboticists aim to build some 'sense organs' (restricted vision and hearing, and recently olfaction), but typically not to teach their robots English. (On natural-language-competent robots, see Shapiro, 2006.) Instead the goal is to give the machines developmentally-earlier communicative abilities, including facial expression and head-and-neck gesture. This sub-field of AI—constructing robots with a 'face' and 'face robots' (machines consisting only of a 'head' and 'face'), and giving them 'facial expressions'—will here be called *face AI*.

A face robot may be quasi-realistic, constructed to resemble a human face and head—for example, Fumio Hara and Hiroshi Kobayashi's series of face robots (Hara 2004; Hara and Kobayashi, 1997) and Chyi-Yeu Lin's face robot singer (Lin et al., 2011) (Figure 1)[1]—or it may be a caricature. (On how people perceive robots with different appearances, see e.g. Lee, Lau, and Hong, 2011.)

[Figure 1 here]

The canonical expressive interactive face robot is Cynthia Breazeal's (now retired) Kismet; this machine has limited movement, visual and auditory systems, and prosodic 'vocalization' (Breazeal, 2002, 2003c) (Figure 2). I shall take Kismet as a paradigm example of an expressive face robot. It is designed to provide 'emotional feedback' and to convey 'intentionality' through facial expressions and behaviour (Breazeal, 2001; Breazeal and Fitzpatrick, 2000). According to Breazeal, 'The robot is able to show expressions analogous to anger, fatigue, fear, disgust, excitement, happiness, interest, sadness, and surprise' (Breazeal and Scassellati, 2000). Her aim is to 'build robots to engage in meaningful social exchanges with humans', and 'the key task for [Kismet] is to apply various communication skills acquired during social exchanges' (Breazeal, 1998a). Kismet, Breazeal claims, 'doesn't engage in adult-level discourse, but its face serves many of these functions at a simpler, pre-linguistic level' (2002, p. 157).

[Figure 2 here]

A social robot like Kismet is designed to be a 'believable creature' (believability is especially important in socially assistive robotics—see Tapus, Matarić, and Scassellati, 2007).[2] Kismet has caricature child-like 'features', to prompt infant-carer behaviour in the human observer (see Section 5). The robot's face responds 'in a timely manner to the person who engages it as well as to other events in the environment', Breazeal says (2002, p. 157). For example, if Kismet 'finds something like a colourful block, it will look at it with a look of happiness because it has found what it wanted. If you play with the robot nicely with the toys, it smiles'.[3] Kismet's 'expressions' are intended to be readable; an untrained human observer should be able to predict and explain the robot's behaviour, by making inferences about the 'emotions' ('goals', 'needs', and so on) manifested in the robot's 'facial' displays (Breazeal, 1999). The aim is that the observer respond to these displays as if to ordinary human social signals, with the result that the displays constrain the human's behaviour in a manner that suits the robot (e.g. Breazeal, 2003a, 2003b). Human observers typically respond to Kismet in exactly this way. For example, one observer 'intentionally put his face very close to the robot's face … The robot withdrew while displaying full annoyance in both face and voice. The subject immediately pushed backwards, rolling the chair across the floor to put about an additional three feet between himself and the robot, and promptly apologized to the robot' (Breazeal and Fitzpatrick, 2000). Some observers even mirror Kismet's behaviour, as they might another human being's behaviour (Breazeal, 2001).

Can a robot *have* emotions? This is currently a much-discussed question in AI (see e.g. Arbib and Fellous, 2004; Adolphs, 2005).[4] Breazeal says that Kismet's 'emotions' are 'quite different' from emotions in humans (Breazeal, 1998a; see also Breazeal, 1999). Kismet's 'emotions' are designed to be 'rough analogs' of human emotions (Breazeal, 1998a). Three similarities between robot and human are implicit in the literature on Kismet. The robot's *system architecture* is influenced by theories in developmental psychology and emotion theory: its 'motivation system' has 'drive' and 'emotion' subsystems that are inspired by work in human ethology. This architecture guides Kismet's interaction with humans—the robot's 'expressions' are designed to evoke responses in humans that conform to its 'goals' and 'drives'. The robot's behaviour is *responsive to the social environment* in ways similar to human social behaviour. For example, given the appropriate states of the 'drive' system, if a human observer 'engages the robot in face-to-face contact … the robot displays *happiness*'; if the slinky toy is removed, 'an expression of *sadness* returns to the robot's face' (Breazeal and Scassellati, 2000).[5] The configurations of the robot's 'facial features' and 'body postures' also copy a human's *expressive behaviours*; 'The robot's facial features move analogously to how humans adjust their facial features to express different emotions', Breazeal claims (1998a).

Typically Breazeal uses scare-quotes or other notational devices when describing the states of Kismet's 'emotion' system. She remarks, for example, that the robot 'responds with an expression of "happiness"', and that 'an expression of "anger" is blended with the intensifying look of "fear"' (Breazeal and Scassellati, 2000, pp. 21, 22). In contrast, Breazeal and other researchers ascribe expressive behaviours to the robot literally and without qualification—thus Kismet 'smiles' (Breazeal, 2000). Kismet is also said, without scare-quotes, to have a 'smile on [its] face', a 'sorrowful expression', a 'fearful expression', a 'happy and interested expression', a 'contented smile', a 'big grin', and a 'frown' (Breazeal, 2001, pp. 584-8; Breazeal and Scassellati, 2001; Breazeal 2000). However, *does* the robot smile, or is this misplaced anthropomorphism?

3. Anatomical Versus Embedded Smiling

According to Wittgenstein, a person who is shown this drawing:

and asked to describe what he or she sees would immediately say 'A face' (PI II, p. 204). This is simply the 'best description I can give of what was shewn me for a moment' (ibid., p. 204). Moreover, anyone shown the drawing 'will be able straight away to reply to such questions as, "Is it male or female?", "Smiling or sad?", etc.' (BB, p. 163). But we are not under 'the delusion of seeing a "real" face', Wittgenstein said (ibid., p. 163). So how can a *drawing* be smiling or sad?

Wittgenstein used the idea of a '*purely* visual concept'—this is a concept that is used 'purely to describe the structure of what is perceived' (LWP I §§736, 739). Applied to a drawing like the one above, the word 'sad', he said, 'characterizes the grouping of lines in a circle' (PI II, p. 209). It is in this sense too that a drawing can be 'smiling'. The purely visual concept of smiling refers simply to, or is based on, the anatomy of the human face; the 'grouping of lines in a circle' of a 'smiling' drawing corresponds to the facial configuration of the smiling human being. I shall call smiling in the purely visual sense *anatomical smiling* (and likewise for frowning, crying, and so on). In this sense a human smiles when only going through the motions of smiling—for example, in response to the orthodontist's instruction to 'smile into the mirror'. In this sense too an emoticon smiles—and Kismet smiles (and also has an angry or sorrowful expression).

According to Wittgenstein, the word 'sad' can in addition be used in a sense that has '*more* than purely visual reference', where it has 'a different, though related, meaning' (PI II, p. 209). Applied to a human being, he said, the concept of sadness is not a purely visual concept—as is shown by the fact that 'sadness I can also hear in his *voice* as much as I can see it in his face' (LWP II §755). Other concepts that might be thought purely sensory likewise contain more. For example, *wailing*: we can use the word 'wailing' to 'describe what is purely acoustical', Wittgenstein said. 'But the truth of the matter is: "Wailing" is not a purely acoustical concept' (LWP I §748). Used of a human being, perhaps the word indicates distress as much as sheer sound.

The word 'smiling' too has a sense that is not purely visual—Wittgenstein remarked that we can speak with 'a *smiling* tone of voice' (LWP I §39). In this more-than-purely-visual sense, what does 'smiling' mean? Wittgenstein said, 'A friendly mouth, friendly eyes, the wagging of a dog's tail are primary symbols of friendliness: they are parts of the phenomena that are called friendliness' (PG, p. 28). According to Wittgenstein, what a behaviour symbolizes is determined by the 'form of life' in which the behaviour occurs; he said, for example, that 'hope, belief, etc. [are] embedded in human life, in all of the situations and reactions which constitute human life' (RPP II §16). For this reason I shall call smiling in the more-than-purely-visual (or symbolic) sense *embedded smiling* (see further Section 4).

Wittgenstein called the drawing above a 'picture-face' (i.e. a picture of a face). He said, 'In some respects I stand towards [the picture-face] as I do towards a human face. I can study its expression, can react to it as to the expression of the human face. A child can talk to picture-men or picture-animals, can treat them as it treats dolls' (PI II, p. 194). This is to anthropomorphize a mere drawing, but only in play. The situation is different with Kismet, however. The descriptions of the robot in the AI literature do not distinguish the anatomical from the embedded sense of 'smile' ('frown', and so on). Developmental robotics may, in consequence, appear to have achieved the goal of building a robot with early communicative abilities—a machine, that is, that smiles, frowns, and so on in the embedded sense—*just because* Kismet 'smiles' in the anatomical sense. Such anthropomorphizing is not harmless.

On the other hand, perhaps Kismet's 'emotion' system and situated behaviour *are* sufficient for the robot's motor acts and 'facial' displays to be genuine expressive behaviour. What does embedded smiling require, and does Kismet smile in this sense?

4. The Wittgensteinian Challenge to 'Strong' Face AI

Applied to face robots, the distinction between anatomical smiling and embedded smiling evokes John Searle's famous distinction between 'weak AI' and 'strong AI' (Searle, 1980). As Searle uses these terms, weak AI is concerned to *simulate* human cognition in a computer; the overall aim is to build machines that, as Marvin Minsky said, 'do things that would require intelligence if done by men' (Minsky, 1968, p. v). Strong AI, in contrast, is concerned to *duplicate* human cognition in a computer; the overarching goal is to build, as John Haugeland said, 'the genuine article: *machines with minds,* in the full and literal sense' (Haugeland, 1985, p. 2). The objective of (what I shall call) *weak face AI* is to build machines that smile (frown, grin, and so on) in the anatomical sense. The objective of *strong face AI* is to build machines that smile in the embedded sense.

For Searle, duplicating human cognition in a computer requires that the machine have the 'awareness' or 'mental life' that humans possess by virtue of their biology; in his view, '[o]nly a being that could have conscious intentional states could have intentional states at all' (1980, pp. 454, 452; 1992, p. 132). Wittgenstein's approach to cognition is very different; in place of (what he called[6]) 'gaseous' conscious states, he emphasized the importance of behaviour, history, and environment (see Proudfoot, 2004a, 2004b; on similarities and differences between Wittgenstein's remarks and Searle's Chinese room argument, see Proudfoot, 2002). His account of facial expression provides (at least part of) the conceptual requirements for strong face AI.

Behaviour's subtle shades

Wittgenstein said, '"Facial expression" exists only within a play of the features' (LWP I §766). The conceptual distinction between *anatomy* and *expression* requires that facial expressions (and gestures and postures) *vary*:

> Suppose someone had always seen faces with only *one* expression, say a smile. And now, for the first time, he sees a face changing its expression. Couldn't we say here that he hadn't noticed a facial expression until now? Not until the change took place was the expression meaningful; earlier it was simply part of the anatomy of the face. (RPP II §356)

'Smiling' is 'our name for an expression in a normal play of expressions', Wittgenstein said (Z §527). *Smiling* is located in a particular space of facial expressions, just as C# belongs in a particular musical scale.

According to Wittgenstein, a smile must also be *mobile*:

> [I]f one were trying to imagine a facial expression not susceptible of gradual and subtle alterations; but which had, say, just five positions; when it changed it would snap straight from one to another. Would this fixed smile really be a smile? And why not?—I might not be able to react as I do to a smile. Maybe it would not make me smile myself. (RPP II §614)

'Variability and irregularity are essential to a friendly expression. Irregularity is part of its physiognomy', he said (RPP II §615). (Interestingly, smiles (in adults) are typically irregular, with greater intensity on the left side of the face—see Nagy, 2012.) Other expressions too must be mobile—would we, Wittgenstein asked, say that a person is in *pain* 'if he always produced exactly the same suffering expression?' (LWP II, p. 67). Likewise grief—'The facial features characteristic of grief … are not more meaningful than their mobility' (RPP II §627). This is an example of the 'importance we attach to the subtle shades of behaviour', Wittgenstein said (ibid., §616). It is in this sense that a 'smiling mouth *smiles* only in a human face'.

The claim that variability and mobility are essential to facial expressions is an empirical hypothesis about how human beings actually react to faces, which can be tested. But it is also a conceptual claim that is implied by Wittgenstein's broader philosophical psychology. In his view, the 'human body is the best picture of the human soul' (PI II, p. 178). The 'soul' that can be seen in a face is just an *aspect* of that face—'And if the play of *expression* develops, then indeed I can say that a soul, something *inner*, is developing', Wittgenstein said (LWP I §947; see further Section 6). This approach to the mind has the consequence that, without the normal play of mobile expressions, 'behaviour would be to us as something completely different' (RPP II §627). Without 'the subtle shades' of behaviour, there is no *soul*, no *inner states*—and no *smiling* in anything but the anatomical sense. There is only 'machine' behaviour.[7]

Although Wittgenstein rejected dualism, his philosophical psychology generates a tough requirement for Kismet's creators. If the robot is to be said to smile in the embedded sense, it must have a normal range and play of human facial expressions and a mobile 'smile'. According to Breazeal, Kismet's 'facial features move analogously to how humans adjust their facial features to express different emotions' (see Section 2). Certainly, the robot's 'features' do move between a number of 'expressions', and the transition from one expression (or posture) to another is smooth and in real time. However, like Lin's face robot singer and other face robots such as ROMAN (Berns and Hirth, 2006), Kismet hardly has a 'normal play' of expressions, and its 'expressions' are *not* variable or irregular—it behaves 'mechanically' (see Section 5). From a Wittgensteinian perspective, Kismet does *not* smile (in the embedded sense).

But is this merely a technological problem? Researchers in human-robot interaction whose aim is to design believable anthropomorphic robots can be seen as trying to capture the 'subtle shades' of human behaviour (see e.g. Lee, Lau, and Hong, 2011). Let us imagine, then, that advances in engineering make possible a descendant of Kismet that produces (anatomical) 'smiles' as part of a normal play of mobile expressions.[8] Would this suffice for strong face AI?

When situated ain't situated

Smiling is a sophisticated behaviour, with multiple meanings. As Wittgenstein pointed out, I can even say "I'm afraid' with a smile (LWP I §21). (On the 'smile of pain', see Kunz, Prkachin, and Lautenbacher, 2009.) Facial expressions work as 'signals', he said, and he sometimes spoke of 'smile-signs' and 'frown-signs' (LPP, p. 283). AI researchers aiming to build robots that actually have emotions focus on *feelings* (e.g.

Adolphs, 2005). However, the difference between an anatomical ‘smile’ and an embedded *smile* is not that the latter is accompanied (or caused) by a feeling of happiness (or some other emotion), since such psychological accompaniments are insufficient to give *meaning* to facial displays.[9] In Wittgenstein’s view, whatever the biological origins and underlying neural mechanisms of smiling, human beings are *trained* to smile, and the significance of smiling derives from this training (RPP I §131).

What makes a ‘smile’ a *smile* is its role in a ‘language game’ or ‘form of life’. This is an example of a general phenomenon—according to Wittgenstein, one can move chess pieces (in accordance with the rules of chess) yet not be *playing chess*, or make ‘Chinese noises’ yet not be *speaking Chinese* (LPP, p. 55), or exhibit some of the expressive behaviour associated with grief yet not be *grieving*:

> ‘Grief’ describes a pattern which recurs, with different variations, in the weave of our life. If a man’s bodily expression of sorrow and of joy alternated, say with the ticking of a clock, here we should not have the characteristic formation of the pattern of sorrow or of the pattern of joy. (PI II, p. 174)

In addition, the same ‘smile’ can be embedded in different behaviour and so acquire different meanings; a ‘grin of friendship and grin of rage may be visually similar, but the consequences are different’, Wittgenstein said (LPP, p. 39). A ‘smile’ takes its meaning from its context:

> I see a picture which represents a smiling face. What do I do if I take the smile now as a kind one, now as malicious? Don’t I often imagine it with a spatial and temporal context which is one either of kindness or malice? Thus I might supply the picture with the fancy that the smiler was smiling down on a child at play, or again on the suffering of an enemy. (PI §539)

For Wittgenstein, expressive behaviour includes not only ‘the play of facial expression and the gestures’ but also ‘the surrounding, so to speak the occasion of this expression’ (RPP I §129). We identify an action ‘according to its background within human life’; this background—‘the whole hurly-burly’ of human actions—‘determines our judgment, our concepts, and our reactions’ (RPP II §624-9). Thus we see a facial display as a *smile* or hear a vocalisation as a cry of *pain* by locating it within what Wittgenstein called ‘the bustle of life’ (ibid., §625). The concept of smiling, like our other concepts, ‘points to something within *this* bustle’ (ibid., §625). A ‘smiling mouth’ *smiles* only when located in this bustle.

Social roboticists recognize that the meaning of a facial expression is context-dependent (see e.g. Hara and Kobayashi, 1997, p. 586). According to Breazeal, Kismet's 'facial expression' *is* a 'signal' to the human observer—the robot's 'sorrowful expression', for example, is 'intended to elicit attentive acts from the human' (2001, pp. 584-5). And the robot's behaviour *is* socially situated; for example, Kismet produces 'smiles' when, after a lack of stimuli, a human being comes into sight—just as a human might do. Nevertheless, from a Wittgensteinian perspective, Kismet's 'expressions' have merely minimal 'surroundings' (to use Wittgenstein's term). This can be seen simply by asking: what *sort* of smile is the robot's 'smile'—friendly, kindly, or malicious? Kismet's 'smile' is none of these, since the robot does not (in fact) have any of the behaviours that are associated with friendliness, or kindness, or malice, and that turn a 'smile' into a *smile*. Likewise for Kismet's 'sad' expression—is the robot grieving, depressed, or sulking? The answer is again: none of these. The consequence is that, contrary to the roboticists' claims, the robot does not *smile*, or have a *sorrowful* expression.

This is (what I shall call) the *embedding problem* for strong face AI. Unless a robot's behaviour is appropriately embedded, to say that the robot *smiles* is misplaced anthropomorphism. We could introduce a technical concept—*smiling**—for which embedding in the 'bustle of life' is unnecessary; but this would not suffice for strong face AI. Even if a future Kismet were to have much greater variability and irregularity of 'facial' displays, strong face AI requires that the robot's displays be smiling in 'the full and literal sense', not merely smiling*.

5. Baby Smiles

People interact with Kismet in some respects as if the robot were an infant, for example speaking to it with the overdone prosody typical of 'motherese' (Breazeal and Aryananda, 2002). Social robots such as Kismet are modelled on human infants. The goal of developmental roboticists is that their robots, from an initial state of mere 'motor babbling' and aided by scaffolding from humans, learn early human social behaviours in a way that is biologically plausible. The researchers' vision is of 'a machine that can learn incrementally, directly from human observers, in the same ways that humans learn from each other' (Scassellati et al., 2006, p. 41; see Shic and Scassellati, 2007). It is in this sense that Kismet is a 'child-machine'.

According to Breazeal, Kismet is constructed 'to receive and send human-like social cues to a caregiver, who can … shape its experiences as a parent would for a child' (Breazeal and Fitzpatrick, 2000, p. 2). A human infant, Breazeal says, naturally 'displays a wide assortment of emotive cues during early face to face exchanges with his mother such as coos, smiles, waves, and kicks'. The mother interprets this behaviour 'as meaningful responses to her mothering and as indications of his internal state'. The infant 'does not know the significance' of his behaviour for his mother, but learns the actions that produce specific responses from her; over time mother and infant 'converge on specific meanings' for these actions.[10] Kismet is designed to generate 'analogous sorts of social exchanges', which help the robot 'learn the meaning [its] acts have for others' (Breazeal, 1998b, p. 31).

The underlying (apparently Vygotskian) hypothesis about cognitive development is that the combination of 'natural' tendencies, motor acts, and human scaffolding generates intentionality. Thus the combination of endogenous 'smiling' and mother-infant imitation (or mimicry, in the case of the infant) produces voluntary social *smiling*. (On the role of cross-cultural as well as maturational factors in the development of social smiling, see Wörmann et al., 2012, and Anisfeld, 1982.) This hypothesis suggests Wittgenstein's own approach to the mind. Nevertheless, employing his criteria for expressive behaviour, it seems that the human infant does *not* smile. Wittgenstein mocked the idea that a baby might have a normal play of expressions—he said, 'Imagine a newborn child who, of course, couldn't speak, but who had the play of features and gestures of an adult!' (LWP I §945). He also said that the 'newborn child' is not capable of being 'malicious, friendly, or thankful. Thankfulness is only possible if there is already a complicated pattern of behaviour' (ibid, §942). The infant does not have any of the behaviours that are associated with friendliness, or kindness, or malice, and that turn a 'smile' into a *smile.* To describe the infant as *smiling* is to anthropomorphize the child, it seems. This result is in tension with psychological studies suggesting that neonates *do* exhibit social smiles (see Cecchini et al. 2011[11]). It also puts Wittgenstein at odds with the standard view that social smiling emerges by approximately 2 months of age (for full-term infants[12]), since at this age the infant's behaviour is not yet embedded in the 'bustle' of life.

On Wittgenstein's account, it appears, neither the robot nor the baby smiles. Yet infant 'smiling' behaviour is very different from Kismet's behaviour, and is closer to that of a paradigm smiling adult. The infant's 'expressions' have greater variability and irregularity than Kismet's displays; according to Wittgenstein, even a dog is 'more like a human being than a being endowed with a human form, but which behaved "mechanically"' (RPP II §623). Wittgenstein also emphasized that human behaviour is embedded in multiple, complicated contexts, and the infant's 'smile' is embedded in many more contexts than is Kismet's 'smile'.

To recognize these differences between the child and the 'child-machine', Wittgenstein's remarks about neonates quoted above must be tempered. In fact he allowed that an infant could have some expressive behaviour, namely imperative pointing. He said that if a child 'stares at the thing with eyes wide open, reaches out towards it with his hand, and perhaps lets out a cry', this 'can perfectly well be called an *expression* of a desire' (PC, p. 419). (The child's 'desire' is not a 'conscious' desire; without language to provide the *object* of desire, there is only 'perhaps an agitation, an obscure urge, a feeling of tightness in the chest' (ibid., p. 421).) This suggests that an infant's 'smile' *could* be called an expression of pleasure (or friendliness). Wittgenstein also said that 'there is such a thing as "primitive thinking" which is to be described via primitive *behaviour*', and that our language-games are 'an extension of primitive behaviour' (RPP II §205; Z §545). Perhaps the child could be said to have 'primitive' emotions which are expressed via primitive expressive behaviour. As usual, we can follow Wittgenstein's injunction to say 'what you choose, so long as it does not prevent you from seeing the facts' (PI §79).[13]

6. The Smiling Pianola

From Kismet's 'smile' we can deduce information about the robot's internal states. According to Breazeal, 'the robot's outwardly observable behavior must serve as an accurate window to its underlying computational processes'—the states of its 'emotion' and 'drive' systems (2002, p. 10). According to Wittgenstein, in contrast, emotions are *not* inner states. Fright, for example, is not 'something which goes along with the experience of expressing fright'; expressing fright is instead *part* of the fright (LPE, p. 202). A corollary of this is that we 'do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom' (RPP II §570). We simply *see* emotion in the face:

> Look into someone else's face and see the consciousness in it, and also a particular *shade* of consciousness. You see on it, in it, joy, indifference, interest, excitement, dullness etc. (RPP I §927)

For Wittgenstein, a facial expression is a (continuous) *aspect* of a face; to see the aspect we simply '[e]nter into the expression' and let 'the face impress itself' on us (RPP I §1033; BB, p. 165). Seeing an aspect 'presupposes concepts which do not belong to the description of the [object] itself', and depends on 'my knowledge, on my general acquaintance with human behaviour' (RPP I §§1030, 1073). Thus seeing a *smile* requires concepts such as friendliness rather than merely anatomical concepts. Wittgenstein's criterion for whether one *sees* an expression is how one responds to the face. For example, he said 'Whoever senses [sadness in a face] often imitates the face with his own' (LWP I §746)—this is affective mirroring, in current terminology. The significance for social cognition of automatically-occurring facial movements in response to facial expressions is a much-researched area in recent years (see e.g. Schilback et al., 2008, on the neural correlates of spontaneous facial reactions to perceived facial expressions). It is striking that Wittgenstein emphasized the importance, in facial-expression processing, of imitating expressions.

Wittgenstein also anticipated the modern discussion of deficits in facial-expression processing—*mind-blindness*, in current terminology. He remarked, 'One might say of someone that he was blind to the *expression* of a face' (LWP I §763); and also, 'I can very well imagine someone who, while he sees a face extremely accurately, and can, e.g., make an accurate portrait of it, yet doesn't recognize its smiling expression as a smile. I should find it absurd to say that his sight was defective' (RPP I §1103). The mind-blind person reacts differently to emotional expressions, according to Wittgenstein: someone who 'sees a smile and does not know it for a smile, does not understand it as such … mimics it differently' (PI II, p. 198). This is consistent with current research on rapid facial responses to emotional expressions that finds differences between the facial reactions of individuals with and without autism spectrum disorders (ASD) (see e.g. Beall et al., 2008).

On Wittgenstein's approach to mind-reading, the mind-blind individual does not lack a *theory* of mind, or the ability to *simulate* another mind, but instead suffers from *aspect*-blindness. Aspect-blindness, Wittgenstein said, is like 'the lack of a "musical ear"' (PI II, p. 214). His approach might generate alternative ways of understanding and identifying mind-blindness, and alternative therapies for impaired performance in (affective) facial-expression recognition. Again Wittgenstein's anticipation of current psychological theorizing is striking. Gangopadhyay and Schilbach (2012), for example, propose an embodiment-based alternative to theory-based and simulation-based analyses of mind-reading, and claim that we have 'immediate perceptual access' to other minds—this is analogous to Wittgenstein's claim that we can look 'into someone else's face and see the consciousness in it'.

According to Wittgenstein, the mind-blind individual can identify facial anatomy (LWP I §780)—that is to say, non-affect facial processing may be unimpaired. Moreover, the mind-blind individual may be able to infer the affect expressed by a face from the geometry of the face, or from other people's responses to the face:

> Think of our reactions towards a good photograph, towards the facial expression in the photograph. There might be people who at most saw a kind of diagram in a photograph, as we consider a map; from it we can gather various things about the landscape (RPP I §170)

Such people might lack the concept of (say) a *hesitant* facial expression, but 'have a concept which was always applicable where "hesitant" … is' (LWP I §741). The mind-blind person's concept would not be equivalent to the original concept (Wittgenstein said, for example, that a *tender* facial expression cannot be described 'in terms of the distribution of matter in space' (LWP I §954)), and Wittgenstein also suggested that there may be facial expressions for which the mind-blind individual would be unable to construct a co-extensive concept.[14] The idea that a person with ASD might have a concept co-extensive with a typically-developing human's concept is an intriguing hypothesis about the facial-expression processing strategies of individuals with and without ASD.[15] Again it is a conceptual claim that forms the basis of an empirically testable hypothesis.

According to Breazeal, robots such as Kismet must be able to 'perceive and interpret the human's emotive expressions', in order to produce appropriate responses (Breazeal, 1999). Kismet has a face detector, and is biased (given the appropriate states of its 'drive' system) to attend to human faces. Other robots, such as Hara and Kobayashi's Mark II robot, are reported to be capable of 'real-time, automatic recognition of human facial expressions' (Hara, 2004; Hara and Kobayashi, 1997). Their robot 'will greet smiles with smiles, frowns with frowns', Hara says (in Menzel and D'Aluisio, 2000, p. 73). The robot Leonardo is designed to learn to reproduce perceived facial expressions, and on this basis is said to imitate (rather than merely mimic) facial expressions. This is to help to bootstrap 'early forms of emotional understanding for the robot'—Leonardo, it is claimed, learns the 'affective meaning' of facial expressions (Breazeal et al., 2005).

However, Kismet, along with other robots designed to recognize facial expressions, lacks the 'general acquaintance with human behaviour' that (according to Wittgenstein) is required for aspect-perception. It follows that the facial-expression processing strategies of the robot and the (typically-developing) human being must differ—Kismet is mind-blind, it seems. This raises the question whether the facial-expression processing strategies implemented in social robots are in fact more like those of individuals with ASD. Such strategies include, it is hypothesized, perceiving faces as a combination of (context-independent) discrete features, in a systematic, sequential manner, versus perceiving faces as a (context-dependent) 'gestalt' whole (for recent research on this see e.g. Evers et al., 2011).

Given the fundamental differences between a 'smiling' robot such as Kismet and a paradigm smiling human, it is misleading to describe such machines as 'child-machines', or as 'socially intelligent'. But then how are we to characterize social robots? In this regard, Wittgenstein's notion of a *reading-machine* is helpful (see Proudfoot, 2002, 2009). A typical reading-machine is a pianola (or 'playing-machine'), which, Wittgenstein said, translates marks into sounds by 'reading' the pattern of perforations in the pianola roll (PI §157; BB, p. 118). A reading-machine 'reads' only in the sense of 'translating script into sounds, also of writing according to dictation or of copying in writing a page of print, and suchlike; reading in this sense does not involve any such thing as understanding what you read' (BB, p. 119).[16] The differences between Kismet and a paradigm smiling human suggest that Kismet is (what I shall call) a *smiling-machine.* (Of course Kismet also produces 'frowns' and other displays—smiling is simply a representative example.) A smiling-machine smiles only in the sense of producing (anatomical) 'smiles' in response to human-appropriate stimuli. Smiling in this sense does not involve *expressing emotion* or *communicating.*

7. Conclusion

Computer scientist Woody Bledsoe famously described a dream 'filled with the wild excitement of seeing a machine act like a human being', and he spoke of the 'yearning' that AI researchers have 'to make machines act in some fundamental ways like people' (1986, p. 57). For many modern researchers, building a robot that 'smiles' and 'frowns', and has a 'happy expression' or 'sorrowful' expression, is an essential step to realizing this dream.

Wittgenstein, however, demonstrated the unrecognized philosophical complexities in this plan.[17] His claim that a 'smiling mouth *smiles* only in a human face' elliptically expresses his claim that smiling not only requires the 'normal play' of human facial expressions but also must be embedded in complicated forms of human life. Without these 'surroundings', an anatomical 'smile' is simply not a *smile*. Wittgenstein's account of facial expression is a storehouse of fascinating ideas for both psychology and AI. It makes clear just how difficult the tasks are for strong face AI, and more generally for cognitive and developmental robotics.

REFERENCES

Wittgenstein

BB (1965). *The Blue and Brown Books*. New York: Harper.

Z (1970). *Zettel*. Edited by G.E.M. Anscombe and G.H. von Wright; translated by G.E.M. Anscombe. Berkeley and Los Angeles: University of California Press.

PI (1972). *Philosophical Investigations* (Second Edition). Edited by G.E.M. Anscombe and R. Rhees; translated by G.E.M Anscombe. Oxford: Basil Blackwell.

PG (1974). *Philosophical Grammar*. Edited by R. Rhees; translated by A.J.P. Kenny. Oxford: Basil Blackwell.

RPP I (1988). *Remarks on the Philosophy of Psychology, Volume I*. Edited by G.E.M. Anscombe and G.H. von Wright; translated by G.E.M. Anscombe. Chicago: University of Chicago Press.

RPP II (1988). *Remarks on the Philosophy of Psychology, Volume II*. Edited by G.H. von Wright and H. Nyman; translated by C.G. Luckhardt and M.A.E. Aue. Chicago: University of Chicago Press.

LPP (1989). *Wittgenstein's Lectures on Philosophical Psychology 1946-47: notes by P.T. Geach, K.J. Shah, and A.C. Jackson*. Edited by P.T. Geach. Chicago: University of Chicago Press.

LFM (1989). *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge 1939: from the Notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees, & Yorick Smythies.*. Edited by C. Diamond. Chicago: University of Chicago Press.

LW I (1990). *Last Writings on the Philosophy of Psychology, Volume I*. Edited by G.H. von Wright and H. Nyman; translated by C.G. Luckhardt and M.A.E. Aue. Chicago: University of Chicago Press.

LW II (1993). *Last Writings on the Philosophy of Psychology, Volume II*. Edited by G.H. von Wright and H. Nyman; translated by C.G. Luckhardt and M.A.E. Aue. Oxford: Blackwell.

PC Wittgenstein, L. and Waismann, F. (2003). 'Psychological Concepts'. In *The Voices of Wittgenstein: The Vienna Circle*. Edited by G. Baker. Translated by G. Baker, J. Connolly, and V. Politis. London: Routledge.

LPE (1993). Notes for Lectures on 'Private Experience' and 'Sense Data'. In *Philosophical Occasions 1912-1951*. Edited by J. Klagge and A. Nordmann. Indianapolis: Hackett, pp. 202-288.

LSD (1993). The Language of Sense Data and Private Experience: Notes by Rush Rhees. In *Philosophical Occasions 1912-1951*. Edited by J. Klagge and A. Nordmann. Indianapolis: Hackett, pp. 290-367.

Other references

Adolphs, R. (2005). Could a Robot have Emotions? Theoretical Perspectives from Social Cognitive Neuroscience. In M. Arbib and J.-M. Fellous (eds), *Who Needs Emotions: The Brain meets the Robot*. Oxford University Press, pp. 9-28. http://www.cc.gatech.edu/~athomaz/papers/adolphs05.pdf.

Anisfeld, E. (1982). The Onset of Social Smiling in Preterm and Full-term Infants from Two Ethnic Backgrounds. *Infant Behavior and Development* 5: 387-95.

Anscombe, G.E.M. (1958). Pretending. *Proceedings of the Aristotelian Society, Supplementary Volume* 32: 279-94.

Aryananda, L. (2006). Attending to Learn and Learning to Attend for a Social Robot. *HUMANOIDS'06—2006 IEEE-RAS International Conference on Humanoid Robots*. http://www.lucs.lu.se/LUCS/128/Aryananda.pdf.

Arbib, M.A. and Fellous, J-M. (2004). Emotions: from brain to robot. *TRENDS in Cognitive Sciences* 8(12): 554-61.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM* 37(7): 122-5. http://dl.acm.org/citation.cfm?id=176803.

Beall, P.M., Moody, E.J., McIntosh, D.N., Hepburn, S.L., and Reed, C.L. (2008). Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder. *Journal of Experimental Child Psychology* 101: 206-23.

Berns, K. and Hirth, J. (2006). Control of facial expressions of the humanoid robot head ROMAN. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* October 9-15, 2006, Beijing, China, pp. 3119-3124. http://agrosy.informatik.uni-kl.de/fileadmin/Literatur/Berns06.pdf.

Bledsoe, W. (1986). I Had A Dream: AAAI Presidential Address, 19 August 1985. *AI Magazine* 7(1): 57-61. http://www.aaai.org/ojs/index.php/aimagazine/article/view/528/464.

Breazeal, C. (1998a). Regulating Human-Robot Interaction using ‘emotions’, ‘drives’ and facial expressions. *Proceedings of 1998 Autonomous Agents workshop, Agents in Interaction -- Acquiring Competence Through Imitation*, Minneapolis, MO, pp. 14-21. http://www.ai.mit.edu/projects/cog/publications.html.

Breazeal, C. (1998b). Early Experiments using Motivations to Regulate Human-Robot Interaction. AAAI Technical Report FS-98-03. http://www.aaai.org/Papers/Symposia/Fall/1998/FS-98-03/FS98-03-007.pdf.

Breazeal, C. (1999). Robot in Society: Friend or Appliance? In *Agents99 workshop on emotion-based agent architectures*, Seattle, WA. 18-26. http://www.ai.mit.edu/projects/sociable/publications.html.

Breazeal, C. (2000). *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. Dissertation submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Doctor Of Science at the Massachusetts Institute of Technology, May 2000. http://groups.csail.mit.edu/lbr/mars/pubs/phd.pdf.

Breazeal, C. (2001). Affective Interaction between Humans and Robots. In J. Kelemen and P. Sosik (eds), *ECAL 2001, LNAI 2159*. Berlin: Springer-Verlag, pp. 582-91.

Breazeal, C. (2002). *Designing Social Robots*. Cambridge, Mass: MIT Press.

Breazeal, C. (2003a). Toward sociable robots. *Robotics and Autonomous Systems* 42: 167-75.

Breazeal, C. (2003b). Emotion and social humanoid robots. *International Journal of Human-Computer Studies* 59: 109-15.

Breazeal, C. (2003c). Emotive qualities in lip-synchronized robot speech. *Advanced Robotics* 17(2): 97-113.

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., and Blumberg, B. (2005). Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots. *Artificial Life* 11(1-2): 31-62.

Breazeal, C. (2006). Human-Robot Partnership. *IEEE Intelligent Systems* 21(4): 79-81.

Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of The Royal Society, Part B* 364: 3527-38.

Breazeal, C. and Aryananda, L. (2002). Recognizing affective intent in robot directed speech. *Autonomous Robots* 12(1): 83-104.

Breazeal, C. and Fitzpatrick, P. (2000). That Certain Look: Social Amplification of Animate Vision. *Proceedings of the AAAI Fall Symposium, Socially Intelligent Agents—The Human in the Loop.* http://people.csail.mit.edu/paulfitz/publications.html.

Breazeal, C., Gray, J., and Berlin, M. (2009). An Embodied Cognition Approach to Mindreading Skills for Socially Intelligent Robots. *The International Journal of Robotics Research* 28: 656-80.

Breazeal, C. and Scassellati, B. (2000). Infant-like Social Interactions Between a Robot and a Human Caretaker. *Adaptive Behavior* 8(1): 49-74. http://www.ai.mit.edu/projects/ntt/projects/NTT9904-01/documents/kismet.pdf.

Breazeal, C. and Scassellati, B. (2001). Challenges in Building Robots That Imitate People. In K. Dautenhahn and C. Nehaniv (eds), *Imitation in Animals and Artifacts*. Cambridge, Mass.: MIT Press. www.ai.mit.edu/projects/humanoid-robotics-group/publications.html.

Breazeal, C. and Velásquez, J. (1998). Toward Teaching a Robot 'Infant' using Emotive Communication Acts. *Proceedings of 1998 Symposium of Adaptive Behavior, Workshop on Socially Situated Intelligence*. http://www.ai.mit.edu/projects/sociable/publications.html.

Brooks, R. A. (1999). *Cambrian Intelligence: The Early History of the New AI.* Cambridge, Mass.: MIT Press.

Brooks, R. A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M.M. (1999). The Cog Project: Building a Humanoid Robot. In C.L. Nehaniv (ed.), *Computation for Metaphor, Analogy, and Agents: Lecture Notes in Artificial Intelligence 1562*. Berlin: Springer, pp. 52-87.

Cecchini, M., Baroni, E., Di Vito, C., and Lai, C. (2011). Smiling in newborns during communicative wake and active sleep. *Infant Behavior and Development* 34: 417-23.

Cecchini, M., Lai, C., and Langher, V. (2007). Communication and crying in newborns. *Infant Behavior and Development* 30: 655-65.

Dautenham, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of The Royal Society, Part B* 362: 679-704.

Evers, K., Noens, I., Steyaert, J., and Wagemans, J. (2011). Combining strengths and weaknesses in visual perception of children with an autism spectrum disorder: Perceptual matching of facial expressions. *Research in Autism Spectrum Disorders* 5: 1327-42.

Fasola, J. and Matarić, M.J. (2012). Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adults. To appear in T. Kanade (ed.), *Proceedings of the IEEE*, *Special Issue on Quality of Life Technology*, http://cres.usc.edu/pubdb_html/files_upload/763.pdf.

Feil-Seifer, D. and Matarić, M.J. (2011). Automated Detection and Classification of Positive vs. Negative Robot Interactions With Children With Autism Using Distance-Based Features. *Proceedings of the International Conference on Human-Robot Interaction* (HRI'11*)*, March 6–9, 2011, Lausanne, Switzerland, pp. 323-30. http://cres.usc.edu/pubdb_html/files_upload/697.pdf.

Fong, T., Nourbakshsh, I., and Dautenham, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems* 42: 143–66.

Gangopadhyay, N. and Schilbach, L. (2012). Seeing minds: A neurophilosophical investigation of the role of perception-action coupling in social perception. *Social Neuroscience* 7(4): 410-23.

Giullian, N., Ricks, D., Atherton, A., Colton, M., Goodrich, M., and Brinton, B. (2010). Detailed Requirements for Robots in Autism Therapy. *2010 IEEE International Conference on Systems, Man, and Cybernetics*, Istanbul, Turkey, October 10-13, 2010. http://faculty.cs.byu.edu/~mike/mikeg/papers/SMC_Detailed_Requirements.pdf.

Gold, K. and Scassellati, B. (2007). A Bayesian robot that distinguishes 'Self' from 'Other'. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society (CogSci2007)*. Nashville, Tennessee. http://cs-www.cs.yale.edu/homes/scaz/papers/Gold-CogSci-07.pdf.

Hara, F. (2004). Artificial Emotion of Face Robot through Learning in Communicative Interactions with Human. *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, Kurashiki, Okayama Japan, September 20-22. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01374712.

Hara, F. and Endo, K. (2000). Dynamic control of lip-configuration of a mouth robot for Japanese vowels. *Robotics and Autonomous Systems* 31: 161–9.
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=537987

Hara, F. and Kobayashi, H. (1997). State-of-the art in component technology for an animated face robot—its component technology development for interactive communication with humans. *Advanced Robotics* 11(6): 585-604.

Haugeland, J. (1985). *Artificial Intelligence: The Very ldea.* Cambridge, Mass.: Bradford Books.

Kim, E.S. and Scassellati, B. (2007). Learning to Refine Behavior Using Prosodic Feedback. *Proceedings of the 6th IEEE International Conference on Development and Learning* (ICDL 2007), London, England, July 2007. http://www.cs.yale.edu/homes/scaz/papers/Kim-ICDL07.pdf.

Kozima, H., Nakagawa, C., and Yano, H. (2005). Using robots for the study of human social development, *AAAI Spring Symposium on Developmental Robotics (DevRob-2005)*, Palo Alto, CA, 22 March 2005, pp. 111-14. http://mainline.brynmawr.edu/DevRob05/schedule/papers/kozima.pdf.

Kunz, M., Prkachin, K., and Lautenbacher, S. (2009). The smile of pain. *PAIN* 145: 273-5.

Lee, S., Lau, I.Y., and Hong, Y. (2011). Effects of Appearance and Functions on Likability and Perceived Occupational Suitability of Robots. *Journal of Cognitive Engineering and Decision Making* 5(2): 232-50.

Lin, C.-Y., Cheng, L.-C., Tseng, C.-K., Gu, H.-Y., Chung, K-L., Fahn, C-S., Lu, K.-J., and Chang, C.-C. (2011). A face robot for autonomous simplified musical notation reading and singing. *Robotics and Autonomous Systems* 59: 943–53.

Menzel, P. and D'Aluiso, F. (2000) *Robosapiens: Evolution of a New Species*. Cambridge, Mass: MIT Press.

Minsky, M. L. (1968). *Semantic Information Processing*. Cambridge, Mass.: MIT Press.

Nagy, E. (2012). From symmetry to asymmetry? The development of smile. *Cortex* 48: 1064-7.

Parisi, D. and Petrosino, G. (2010). Robots that *have* emotions. *Adaptive Behavior* 18: 453-69.

Proudfoot, D. (2002). Wittgenstein's Anticipation of the Chinese Room. In J. M. Preston and M. A. Bishop (eds.), *Views Into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford: Oxford University Press, pp. 167–80.

Proudfoot, D. (2004a). Robots and Rule-following. In C. Teuscher (ed.), *Alan Turing: Life and Legacy of a Great Thinker*. Berlin: Springer-Verlag, pp. 359–79.

Proudfoot, D. (2004b). The implications of an externalist theory of rule-following behaviour for robot cognition. *Minds and Machines* 14: 283–308.

Proudfoot, D. (2009). Meaning and mind: Wittgenstein's relevance for the 'Does Language Shape Thought?' debate. *New Ideas in Psychology* 27: 163–83.

Proudfoot, D. (2011). Anthropomorphism and AI: Turing's Much Misunderstood Imitation Game. *Artificial Intelligence* 175: 950-7.

Scassellatti, B. (2000). How Developmental Psychology and Robotics Complement Each Other. *NSF/DARPA Workshop on Development and Learning*, Michigan State University, Lansing, MI, June 2000. http://www.cs.yale.edu/homes/scaz/papers/WDL2000.pdf.

Scassellatti, B. (2001). Investigating Models of Social Development Using a Humanoid Robot. In B. Webb and T. Consi (eds), *Biorobotics*. Cambridge, Mass.: MIT Press. http://www.cs.yale.edu/homes/scaz/papers/cogbio.pdf.

Scassellatti, B. (2002). Theory of Mind for a Humanoid Robot. *Autonomous Robots* 12: 13–24.

Scassellatti, B. (2005). How social robots will help us to diagnose, treat, and understand autism. *12th International Symposium of Robotics Research (ISRR)*, San Francisco, Calif., Oct. 2005. http://www.cs.yale.edu/homes/scaz/papers/Scassellati-ISRR-05-final.pdf.

Scassellati, B., Crick, C., Gold, K., Kim, E., Shic, F., and Sun, G. (2006). Social Development. *IEEE Computational Intelligence Magazine*, August 2006, pp. 41-47. http://www.cs.yale.edu/homes/scaz/papers/Scaz-Computational-Intelligence.pdf.

Schilbach, L. Eickhoff, S.B., Mojzisch, A. and Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience* 3(1): 37-50.

Searle, J.R. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences* 3: 417-56.

Searle, J.R. (1992). *The Rediscovery of the Mind.* Cambridge, Mass.: MIT Press.

Shapiro, S.C. (2006). Natural-Language-Competent Robots. *IEEE Intelligent Systems* 21(4): 76-7.

Shic, B. and Scassellati, B. (2007). Pitfalls in The Modeling of Developmental Systems. *International Journal of Humanoid Robotics* 4(2): 435-54. http://www.cs.yale.edu/homes/scaz/papers/Shic-IJHR2007.pdf.

Shic, B., Jones, W., Klin, A., and Scassellati, B. (2006). Swimming in the Underlying Stream: Computational Models of Gaze in a Comparative Behavioral Analysis of Autism. *Cognitive Science*, Vancouver, 2006, pp. 780-5. http://www.cs.yale.edu/homes/scaz/papers/Shic-CogSci2006.pdf.

Sun, G. and Scassellati, B. (2005). A Fast And Efficient Model For Learning To Reach. *International Journal of Humanoid Robotics* 2(4): 391-413. http://www.cs.yale.edu/homes/scaz/papers/IJHR-05-Sun.pdf.

Tapus, A., Matarić, M.J., and Scassellati, B. (2007). The Grand Challenges in Socially Assistive Robotics. *IEEE Robotics And Automation Magazine* 4(1): 35-42. http://www.cs.yale.edu/homes/scaz/papers/Tapus-RAM2007.pdf.

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind* 59: 433-60.

Wade, E., Parnandi, A., Mead, R., and Matarić, M.J. (2012). Socially Assistive Robotics for Guiding Motor Task Practice. *Journal of Behavioral Robotics* 2(4): 218-27. http://cres.usc.edu/pubdb_html/files_upload/764.Paladyn.2.2011.pdf.

Wörmann, V., Holodynski, M., Kärtner, J., and Keller, H. (2012). Smiling in newborns during communicative wake and active sleep. *Infant Behavior and Development* 34: 417-23.

---

[1] The researchers building these face robots are concerned more with the mechanics of human facial expression, and modelling *anatomical* smiles, rather than giving the robot a 'motivational' system based upon theories of emotions in humans and other animals (see e.g. Hara and Endo, 2000).

[2] On 'believable creatures' see Bates, 1994.

[3] Cynthia Breazeal, quoted in Menzel and D'Aluisio, 2000, p. 69.

[4] Typically AI researchers deny that their 'emotional' robots actually have emotions. Parisi and Petrosino are an exception; they claim that their simulated robots '*have* emotions because emotions can be shown to play a well-identified function in what they do' (2010, p. 453).

[5] Breazeal distinguishes Kismet's 'emotive expressions', which 'reflec[t] the state of the robot's emotion system' from its 'expressive facial displays', which 'conve[y] social cues during social interactions with people' (2002, p. 158). I shall use the words 'expression' and 'display' interchangeably.

[6] LSD, p. 367.

[7] Wittgenstein said, for example, 'How would a human body have to act so that one would not be inclined to speak of inner and outer human states? Again and again, I think: "like a machine".' (LWP II, p. 66).

[8] The assumption that this is solely an *engineering* matter is harmless here, given the additional challenge for AI of the embedding problem.

[9] Nor are psychological accompaniments *necessary*; non-genuine ('Duchenne') smiles have the same meaning as genuine smiles, even though ex hypothesi the insincere person lacks the supposedly appropriate feeling. In any case Wittgenstein argued that pleasure is not a sensation, and that emotions are not sensations (RPP I §800; RPP II §148).

[10] http://www.ai.mit.edu/projects/kismet-new/kismet.html.

[11] For the related claim that crying in newborns depends, not only on the infant's psychological state, but on the communicative context, see Cecchini et al., 2007.
[12] See Anisfeld 1982.
[13] See Elizabeth Anscombe's article 'Pretending' for this approach to the anthropomorphizing of non-linguistic animals (1958, p. 291).
[14] Using an analogy with musical expression, Wittgenstein said that a *tender* expression in music 'can't even be explained by reference to a paradigm, since there are countless ways in which the same piece may be played with genuine expression' (LWP I §954).
[15] On the computational modelling of gaze, comparing autistic individuals and controls, see e.g. Shic et al., 2006.
[16] A reading-machine can be 'living' (PI §157)—for example, Wittgenstein said that a living calculating-machine can produce proofs of complicated mathematical theorems but cannot figure out 'what change you should get from a shilling for a twopenny bar of chocolate' (LFM, p. 36). Might we then regard the mind-blind *human being* (or even the human infant) as a *living smiling-machine*—a human who produces only (anatomical) 'smiles', and does not express emotion by this behaviour?
[17] This paper was written while I was a Visiting Research Fellow in the Department of Psychology at Georgetown University. I am extremely grateful to the Department, in particular to James Lamiell and Rom Harré, and to Rachel Barr for valuable discussion about imitation in infancy.