# Robot nannies

## Future or fiction?

Egon L. van den Broek
Human-Centered Computing Consultancy, Nijmegen, The Netherlands/
Human Media Interaction, University of Twente, Enschede, The Netherlands/
Karakter, Radboud University Medical Center, Nijmegen, The Netherlands

**Keywords:** robot nannies; artificial intelligence; child development; sensors; context

## 1. Introduction

Most of us will have an idea of what a robot nanny would or should be. Although often omitted, also in scientific literature, it is good to ensure a common ground on key notions. Merriam-Webster, Incorporated (2010) defines a robot simply as "a mechanism guided by automatic controls".[1] The same dictionary defines a nanny as "a child's nurse or caregiver". Note that the former definition is a definition with a wide scope. In contrast, the latter definition identifies two rather concrete functions. Ideally, a nanny embodies both of them; however, in practice this is not often the case. In their article, Sharkey & Sharkey (2010) assume that a nanny is a child's caregiver and so do we. Subsequently, a combination of both definitions provides us a stipulative definition of robot nannies: A mechanism guided by automatic controls which functions as a child's caregiver. However, it is questionable whether this definition fits the idea we had of robot nannies.

Robot nannies can also be considered as a subclass of social robots, i.e. "… the class of robots that people anthropomorphize in order to interact with them" (Breazeal, 2003, p. 167). This is perhaps a more intuitive definition, more close to our idea of robot nannies. Various social robots have been developed in first decade of the 21st century (Breazeal, 2003). If anything, this illustrates the complexity of this challenge, which requires knowledge from a broad range of expertise; see also (Siciliano & Khatib, 2008, Chapter 58). Robot nannies have their own requirements, not in the last place due to its specific group of users: children. Consequently, robot nannies have a big responsibility or should we say their developers have this responsibility? Either way, the development of robust robot nannies is a tremendous challenge.

*Interaction Studies* 11:2 (2010), **274–282**. DOI 10.1075/is.11.2.16van
ISSN 1572–0373 / E-ISSN 1572–0381 

This commentary discusses some of the challenges robot nannies have to overcome, which are only briefly touched on by Sharkey & Sharkey (2010). As a foundation for this discussion, we introduce a general model of a robot nanny, as is shown in Figure 1. This model denotes all relevant aspects of a robot nanny and relates it to the child, the context, and possibly a supervising human nanny. This article discusses, subsequently, (i) the child model, taking into account child development (including interpersonal differences), (ii) the interface, providing robot nannies with sensors, and (iii) the issue of context. Parallels are drawn with other branches of artificial intelligence (AI), which illustrate the problems robot nannies have to face. On the one hand, classic theories will be touched on; on the other hand, state-of-the-art research is included. We end this commentary, with a reflection on AI and general conclusions.

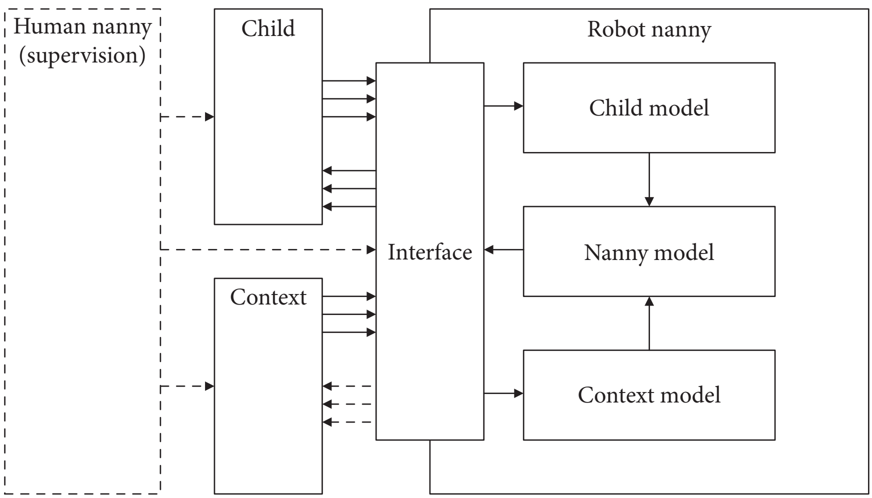


**Figure 1.** Robot nanny's model and its related aspects: child, context, and human nanny as possible supervisor. The robot nanny consists of three models for the child, the nanny, and the context. In addition, the robot nanny has an interface. This interface perceives the child as well as its environment and, subsequently, processes and classifies its input. The lines denote information streams, where dashed lines denote optional information streams

## 2. Child development

Although criticized frequently, Piaget's stage theory is still a generally accepted theoretical framework for child development (Louren¸co & Machado, 1996). Piaget identified four stages (a.k.a. periods): the sensorimotor (0–2 years), the preoperational (2–7 years), the concrete operational (7–11 years), and the formal operational

(11–15 years) period. Each of these periods is described by typical characteristics of child development. Throughout 15 years, children move from simple reflexes to organized behavior, to rigidity of thought, to the application of operations to temporal-spatial representations, and to entities with complete cognitive structures, who act through a tightly organized system of thought – a unified whole. Although this is only a brief sketch and various other theories exist, it illustrates the complexity of child care which robot nannies have to face. Such a theory could serve as foundation for the child model embedded in the robot nanny; see also Figure 1.

Piaget's generic framework does not take into account individual differences at their various stages of development. However, such differences can influence children's behavior significantly; hence, a robot nanny's child model should also take this into account (Marwitz & Stemmler, 1998). For example, differences between children are found in infant attachment, defined by Sharkey & Sharkey (2010) as "the deep emotional connection that an infant forms with his or her primary caregiver, often the mother." (p. 174). For children's development, a good attachment is crucial. As the authors later also state, "Responding appropriately to an infant's cues requires a sensitive and subtle understanding of the infant's needs." (p. 174). This is also known as maternal sensitivity (Donovan, Leavitt, & Balling, 1978), as the robot nanny should have incorporated in its nanny model (see Figure 1). The seminal work of Harlow (cited in Sharkey & Sharkey, 2010) on baby monkeys already showed the importance of attachment for babies. Moreover, it enabled the identification of problems that emerge with insecure attachment. However, how caretakers should behave differs among children and mainly relies on maternal sensitivity (Donovan et al., 1978).

Understanding children requires for instance also that their personality has to be taken into account. The fact that their personality is in development does not diminish the need for differentiation between them. Also cultural aspects should be taken into account, since they can play a significant role in the behavior of a child. A robot nanny should also take into account these issues, otherwise its understanding of children will remain a myth.

Taken in a nutshell, as Sharkey & Sharkey (2010) state: "There are many cues that an adult human uses to understand what answer the child requires and at what level." (p. 175). These are processed through robot nanny's interface; see also Figure 1. This leaves us with the questions on which cues to sense and how to process them.

## 3. Robot nanny's sensors

If anything, robot nannies should sense children's emotions; cf. Donovan et al. (1978) and Siciliano & Khatib (2008, Chapter 58). As Dautenhahn, Bond, Cañamero, & Edmonds (2002, p. 6) stated: "Agents that can recognize a user's emotions, display meaningful emotional expressions, and behave in ways that are perceived as coherent, intentional, responsive, and socially/emotionally appropriate, can make important contributions towards achieving human–computer interaction that is more 'natural', believable, and enjoyable to the human partner." This statement is in line with the general notion that emotions are the missing link in AI and human–robot interaction (Broek et al., 2010; Healey & Picard, 2005; Picard, 1997).

Emotions cannot be ignored; they influence us, be it consciously or unconsciously, in a wide variety of ways (Picard, 1997; Vinciarelli, Pantic, & Bourlard, 2009):

- long term physical well-being; e.g. cardiovascular issues (Frederickson, Manusco, Branigan, & Tugade, 2000) and our immune system (Ader, Cohen, & Felten, 1995);
- physiological reactions/biosignals (Agrawal, Liu, & Sarkar, 2008; Broek et al., 2010; Donovan et al., 1978); e.g. as present in communication;
- cognitive processes; e.g. perceiving, memory, reasoning (Critchley, Elliott, Mathias, & Dolan, 2000); and
- behavior (Vinciarelli et al., 2009); e.g. facial expressions, speech (Gelder, 2009; Zeng, Pantic, Roisman, & Huang, 2009), movements, and touch (Bailenson, Yee, Brave, Merget, & Koslow, 2007; Gelder, 2009; Poppe, 2010).

Children, being in development, are more vulnerable to influences on each of these levels than adults are.

Automated emotion recognition, also known as affective computing, can be realized through the recording of various signals that humans transmit continuously (Broek et al., 2010; Picard, 1997; Vinciarelli et al., 2009) and that can be perceived through a robot nanny's interface; see Figure 1. As Sharkey & Sharkey (2010) mention, the most often employed techniques comprise computer vision, movement analysis, and speech processing. However, each of these techniques, has its drawbacks, in particular in ambulatory settings such as those of robot nannies:

- recording and processing of facial expressions assessed through computer vision techniques suffer from changes in light; e.g. causing a lack of contrast or shadows (Zeng et al., 2009; Gunes & Piccardi, 2009)
- movement analysis requires continuous tracking, which is often not feasible in practice (Gunes & Piccardi, 2009; Poppe, 2010), and
- speech processing (Broek, Schut, Westerink, & Tuinenbreijer, 2009; Zeng et al., 2009) suffers from severe distortions (e.g. environmental noise, multiple voices in parallel).

These concerns limit the feasibility of emotion recognition by robot nannies significantly. Hence, either other signals need to be identified for emotion recognition or emotion recognition in the wild is not possible (Broek et al., 2010).

A possible solution lays in what William James already noted in 1893, humans are *psycho-neuro-physical mechanisms*, who both send and perceive biosignals that can be captured; e.g. electromyography, electrocardiography, and electrodermal activity. These biosignals can also be used to reveal a range of characteristics of people, among which are emotions (Broek et al., 2010; Friedman, 2010; Healey & Picard, 2005; Picard, 1997).

Biosignals have the additional advantage that they are free from social masking (Broek et al., 2010). Moreover, nowadays, they can be measured by non-invasive unobtrusive sensors, which makes them suited for child care (Broek et al., 2010; Thiemjarus & Yang, 2006). So, biosignals could act as an interface between infants and their robot nanny (see also Figure 1, which could provide robot nannies with the information to develop empathic abilities.

However, it should be noted that biosignals, computer vision, movement analysis, and speech processing have some other serious drawbacks. The children differ in how their emotions and the accompanying signals are related (Stemmler & Wacker, 2010). Moreover, their personality can play a significant role in how the emotions are expressed through biosignals (Broek et al., 2009; Stemmler & Wacker, 2010).

Although combinations of either computer vision and speech (Zeng et al., 2009), computer vision and movements/touch (Gunes & Piccardi, 2009), computer vision and biosignals (Bailenson et al., 2008), or speech and biosignals (Broek et al., 2009) are still rare, initial results are promising. Hence, such a triangulation of emotions could possibly be a solution to the problems mentioned (Broek et al., 2010), which make the development of a robot nanny's interface (see Figure 1) very complex. However, as with most research towards affective computing, these studies were conducted in (semi-) controlled environments. One of the rare truly ambulatory studies conducted illustrates the complexities that arise with affective computing in the wild (Healey & Picard, 2005). Hence, it is unknown to what extent they are of use in ambulatory settings such as those in which child care takes place.

More recently, brain-computer interfaces (BCI) have became popular (Gerven et al., 2009; Nijholt et al., 2008). BCI can be considered as advanced, massively parallel biosignal (e.g. EEG) recording and (real-time) processing. BCI could be an interesting interface for robot nannies since significant results have already been reported in neuroscience concerning emotions (Gelder, 2009); e.g. with respect to mirror neurons, as is also mentioned by Sharkey & Sharkey (2010). This new field of research is starting to develop its guidelines (Gerven et al., 2009), and massive investments are being made to ensure future progress. However, so far, BCI has not redeemed its promises and is too obtrusive and noise sensitive for ambulatory applications such as robot nannies.

## 4. Context

Robot nannies are an example of AI and, as such, illustrate both AI's advances and its limitations; cf. Siciliano & Khatib (2008). One of AI's traditional struggles is its battle to deal adequately with context (Avidan & Shamir, 2007; Broek et al., 2010; Thiemjarus & Yang, 2006). This is no different for robot nannies; cf. Figure 1. In this section, we will discuss branches of AI that have common characteristics with the field of robot nannies. Consequently, important lessons can be learned from them.

With striking ease, humans are able to think and act based on incomplete information. As Sharkey & Sharkey (2010, p. 175) state, "A human carer may not get a full and complete understanding of the context of an emotion every time but they will make a good guess with a high hit rate and can then recalculate based on the child's subsequent responses." Despite all research efforts on understanding human multisensory processing (Calvert, Spence, & Stein, 2004), human cognition (Perlovsky, 2007), and new techniques on multimedia processing (Avidan & Shamir, 2007; Milisavljevic, 2009; Vinciarelli et al., 2009), science is still struggling with the low level processing issues. So far, no general fusion paradigm or processing framework has been presented. For specific applications such as robot nannies, tailored solutions are developed (Siciliano & Khatib, 2008, Part F). However, their success is limited and hence, the techniques are not reliable enough to be used for robot nannies.

A decade ago, Emile Aarts (1999) coined Ambient Intelligence (AmI), which has evolved to a field on its own, together with smart sensor networks and their applications (e.g. smart homes for the elderly). The AmI paradigm nicely illustrates the need for advanced multisensor data gathering, processing, and interpretation, as is also the case for robot nannies (Siciliano & Khatib, 2008, Part C). Not the gathering is a problem, but processing (e.g. feature extraction) and interpretation is hard (Avidan & Shamir, 2007; Broek et al., 2009; Poppe, 2010; Vinciarelli et al., 2009). All this should be done in the robot nanny's interface; see Figure 1. AmI studies reveal the problems science and engineering have to face when developing robot nannies.

Although assumed, the explicit notion that we need models, was not been discussed so far. Intelligent Tutoring Systems (Anderson, Boyle, & Reiser, 1985, ITS), a branch of expert systems that was vivid 25 years ago, extensively used models of children (i.e. the students aimed to use the ITS). However, these were tailored towards a specific (sub)domain, where robot nannies would need (more) generic models. This is precisely what ITS failed to deliver. Nevertheless, the choice of a similar approach for robot nannies would be an obvious choice. Consequently, the robot nanny model shares some characteristics with ITS models; cf. Figure 1. However, as ITS research has shown, we have a long way to go before such models could be applied in practice.

## 5. Conclusions

With this discussion on robot nannies, a range of aspects of AI were taken into account; see also Figure 1. Once more, it was illustrated how brittle AI's advances are. In particular, the lack of integration of various paradigms and processing schemes is striking. AI seems to be scattered in itself and knowledge transfer between its subfields is limited. To enable the determination of (lack of) progress, recently, an initiative on benchmarks for human–robot interaction was launched (MacDorman & Kahn, 2007). Specific benchmarks should be developed for robot nannies. Such benchmarks would facilitate the development of a general progress indicator for robot nannies and for AI and robotics in general; cf. Siciliano & Khatib (2008, e.g. Chapter 50).

As Sharkey & Sharkey (2010, p. 177) state, "While it seems unlikely that a robot could show a sufficient level of sensitivity to engender secure attachment, it could be argued that the robot is only be standing in for the mother in the same way as a human nanny stands in. But a poor nanny can also cause emotional or psychological damage to a child.". Regrettably, no other conclusion can be drawn than that good, reliable robot nannies are beyond current state-of-the-art AI. Even with the progress ahead and vast investments being made, it is questionable whether the time will come when robot nannies will take care of children. So, the question remains, will robot nannies be something of the future or will they remain fiction?

## Note

1. For an excellent general introduction in robotics, we refer to Siciliano & Khatib (2008)

## References

Aarts, E., & Appelo, L. (1999). Ambient intelligence: Thuisomgevingen van de toekomst. IT-Monitor, #9, 4–7.

Ader, R., Cohen, N., & Felten, D. (1995). Psychoneuroimmunology: Interactions between the nervous system and the immune system. The Lancet, 345(8942), 99–103.

Agrawal, P., Liu, C., & Sarkar, N. (2008). Interaction between human and robot: An affect-inspired approach. Interaction Studies, 9(2), 230–257.

Anderson, J.R., Boyle, C.F., & Reiser, B.J. (1985). Intelligent Tutoring Systems. Science, 228(4698), 456–462.

Avidan, S., & Shamir, A. (2007). Seam carving for content-aware image resizing. ACM Transactions on Graphics, 26(3), Article No. 10.

Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., et al. (2008). Real-time classification of evoked emotions using facial feature tracking and physiological responses. International Journal of Human-Computer Studies, 66(5), 303–317.

Bailenson, J.N., Yee, N., Brave, S., Merget, D., & Koslow, D. (2007). Virtual interpersonal touch: Expressing and recognizing emotions through haptic devices. Human–Computer Interaction, 22(3), 325–353.

Breazeal, C. (2003). Toward sociable robots. Robotics and Autonomous Systems, 42(3–4), 167–175.

Broek, E.L. van den, Lisý, V., Janssen, J.H., Westerink, J.H.D.M., Schut, M.H., & Tuinenbreijer, K. (2010). Affective man–machine interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe, & H. Gamboa (Eds.), Biomedical engineering systems and technologies: BIOSTEC2009 selected revised papers (Vol. 52, pp. 21–47). Berlin/Heidelberg, Germany: Springer-Verlag.

Broek, E.L. van den, Schut, M.H., Westerink, J.H.D.M., & Tuinenbreijer, K. (2009). Unobtrusive Sensing of Emotions (USE). Journal of Ambient Intelligence and Smart Environments, 1(3), 287–299.

Calvert, G.A., Spence, C., & Stein, B.E. (2004). The handbook of multisensory processes. MIT Press, USA.

Critchley, H.D., Elliott, R., Mathias, C.J., & Dolan, R.J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. The Journal of Neuroscience, 20(8), 3033–3040.

Dautenhahn, K., Bond, A.H., Cañamero, L., & Edmonds, B. (2002). Socially intelligent agents: Creating relationships with computers and robots (Vol. 3). New York, NY, USA: Kluwer Academic Publishers.

Donovan, W.L., Leavitt, L.A., & Balling, J.D. (1978). Maternal physiological response to infant signals. Psychophysiology, 15(1), 68–74.

Frederickson, B.L., Manusco, R.A., Branigan, C., & Tugade, M.M. (2000). The undoing effect of positive emotions. Motivation and Emotion, 24(4), 237–257.

Friedman, B.H. (2010). Feelings and the body: The Jamesian perspective on autonomic specificity of emotion. Biological Psychology, 83(x), [in press].

Gelder, B. de. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3475–3484.

Gerven, M. van, Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., et al. (2009). The brain–computer interface cycle. Journal of Neural Engineering, 6(4), 1–10.

Gunes, H., & Piccardi, M. (2009). Automatic temporal segment detection and affect recognition from face and body display. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 39(1), 64–84.

Healey, J.A., & Picard, R.W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, 6(2), 156–166.

James, W. (1893). Review: La pathologie des emotions by Ch. Féré. The Philosophical Review, 2(3), 333–336.

Lorenço, O., & Machado, A. (1996). In defense of Piaget's theory: A reply to ten common criticisms. Psychological Review, 103(1), 143–164.

MacDorman, K.F., & Kahn, P.H. (2007). Introduction to the special issue on psychological benchmarks of human–robot interaction. Interaction Studies, 8(3), 359–362.

Marwitz, M., & Stemmler, G. (1998). On the status of individual response specificity. Psychophysiology, 35(1), 1–15.

Merriam-Webster, Incorporated. (2010). Merriam-Webster Online. URL: http://www.m-w.com/ [Last accessed on May 5].

Milisavljevic, N. (2009). Sensor and data fusion. Vienna, Austria: In-Tech.

Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C., Millán, J.R., Allison, B., et al. (2008). Brain–computer interfacing for intelligent systems. IEEE Intelligent Systems, 23(3), 72–79.
Perlovsky, L.I. (2007). Cognitive high level information fusion. Information Sciences, 177(10), 2099–2118.
Picard, R.W. (1997). Affective computing. Boston, MA, USA: MIT Press.
Poppe, R. (2010). A survey on vision-based human action recognition. Image and Vision Computing, 28(6), 976–990.
Sharkey, N., & Sharkey, A. (2010). The crying shame of robot nannies: An ethical approach. Interaction Studies, 11, [this issue].
Siciliano, B., & Khatib, O. (2008). Springer handbook of robotics. Berlin/Heidelberg, Germany: Springer-Verlag.
Stemmler, G., & Wacker, J. (2010). Personality, emotion, and individual differences in physiological responses. Biological Psychology, 83(x), [in press].
Thiemjarus, S., & Yang, G.-Z. (2006). Context-aware sensing. In G.-Z. Yang (Ed.), Body sensor networks (pp. 287–331). London, UK: Springer–Verlag London Limited.
Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and Vision Computing, 27(12), 1743–1759.
Zeng, Z., Pantic, M., Roisman, G.I., & Huang, T.S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1), 39–58.

## *Author's address*

Egon L. van den Broek
Human-Centered Computing Consultancy
Moeflonstraat 76
6531 JR Nijmegen, The Netherlands

Email: vandenbroek@acm.org

## *Biographical note*

**Egon L. van den Broek** obtained his MSc (2001) in Artificial Intelligence (AI) and Ph.D. (2005) in Content-Based Image Retrieval and expects to obtain a Ph.D. (2010) in Affective Signal Processing. Currently, he is assistant professor/research coordinator (University of Twente and Radboud University Medical Center, both NL), as well as consultant with Human-Centered Computing Consultancy. In the latter function, he works for various companies, among which Philips Research. Within the fields of AI and human media interaction, his primary research interests are cognitive computer vision and affective computing. He has published 100+ articles, guided 45+ students, holds various awards, is or has been reviewer and/or program committee member of 50+ conferences and journals, and holds 3 patent applications. Additionally, he is editor-in-chief of the Pan Stanford Series in Artificial Intelligence, which will be launched in 2011/2012.