How to make agents that display believable empathy? An ethological approach to empathic behavior (Extended Abstract)

Ádám Miklósi
Eötvös Lóránd University
Pázmány P s 1c
Budapest, 1117 Hungary
+36 1 382 27 79
amiklosi62@gmail.com

ABSTRACT
The plan to engineer "empathic agents" is very ambitious, specifically because many researchers resist attributing such ability to any animal other than humans. Thus it seems to be paradoxical to have emphatic agents but no empathic animals. This review suggests that affective computing may be boosting force for developing a unified approach to the evolution in empathic behaviour in living systems, and the knowledge gained could be utilised for designing machines that produce empathic behaviour which is believable for the human partners.

General Terms
Design, Human Factors, Theory

Keywords
Empathy, animals, evolution, inter-specific

1. INTRODUCTION
The scientific interest in empathic behaviour has a long story in the psychological sciences. Although it was often used as an explanatory term for many aspects of human behaviour, specific research was lacking. Among other factors the so called „cognitive revolution" in psychology facilitated research in this topic, especially by studying the developmental aspects of emphatic behaviour in human children.

As animals (e.g. rats) have been often utilised as models of human behaviour, already in the 60ies researchers demonstrated „empathy-like" behaviour in rats. If a rat had observed a stressful con-specific that was suspended in the air by a harness, it moved to press the bar in order to lower the rat back to ground [9]. Although such laboratory investigations of „animal models" documented many situations when the behaviour of the observer animal could be interpreted as being driven by „empathy", researchers were reluctant to argue for basic human-animal similarities in the underlying mechanisms. Even today many researchers avoid referring to empathy altogether when explaining some social behaviour, or they put the word in quotations.

Based on the arguments put forward by Darwin [2] on the continuity of „mental abilities and emotional expression" in evolution, interest has emerged to look for phylogenetic roots of human empathy in animals (for comparative review see 8).

2. DEFINITION OF EMPATHY
The definition of empathy suffers from problems that are common with terms that are used in everyday situations, and which are associated with specific human abilities. Even if researchers try to be objective, they have difficulties to avoid a human-centred view (anthropocentrism) that is often combined with „unconscious" introspective tendencies. Thus for many researchers empathic ability equals the „capacity for putting oneself in somebody's place". This approach is in many ways analogous to what is attributed to "mind reading". It is not surprising that psychologists prefer to talk about understanding another's emotional state, and refer to unobservable cognitive states when explaining the mechanisms controlling empathic behaviour in humans. This attitude is problematic because it is difficult to utilise such a research agenda in a comparative perspective if one is interested in the evolutionary origin of empathic behaviour.

For example, how can be utilise Hoffman's [3] widely cited definition of empathy (*„any process where the attended perception of the object's state generates a state in the subject that is more applicable to the object's state or situation than to the subject's own prior state or situation")* in the case of animals or especially artificial agents? It would be very difficult to argue for empathy in animals, and researchers would be accused of anthropomorphism, because there is no objective method for the comparison of inter-specific or inter-agent inner states.

In line with this criticism Preston and de Waal [8] use a somewhat extended definition for their „Perception-Action Model" of empathic behaviour. They argue that the *„attended perception of the object's state automatically activates the subject's representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited".* This definition is more useful because it refers no just to states but also to the behaviour (at least on the part of the subject). It is still problematic that in the discussion of empathy researchers move to quickly to the underlying (and unobservable) mental states of the mind and pay much less attention to investigate the mechanisms at the behavioural level. This situation creates often a terminological confusion in the use of categories; especially because researchers have a tendency of re-use value-loaded verbal expressions of human behaviour features (e.g. „sympathy").

In the following we will follow Tinbergen's [10] receipt and look for possible functions of behaviours that might be interpreted as being „empathic". Ideas based on evolutionary considerations will help us in this case.

3. EVOLUTION OF EMPATHIC BEHAVIOUR

Already Darwin attempted to explain the evolutionary origin of empathic behaviour. He and later other argued that such interactions might be very important in the mother-infant relationship [2], especially in mammals in which we find a very intensive and often long-lasting parental care. Empathic behaviour could mutually strengthen this bond and contribute to the survival of the offspring.

Interest in altruistic ("unselfish") behaviour among animals [11] led to the assumption that inclusive fitness and reciprocal altruism could explain the evolution of empathy. In this model empathy is the mechanism, which facilitates the mutual relationship between the interacting partners. Thus this is is an extension of the empathic aspects of mother-infant bond to relatives or even unrelated group members.

More recently, Preston and de Waal [8] argued for an even more general evolutionary function for empathic behaviour. They suggest that the phylogenetic explanation of empathic behaviour can be found in social animals in which the synchronic activities of the group are of vital importance. According to this scenario social animals would be at an advantage to display similar behaviours, that is, if one animal responds with a matching action after having perceived the behaviour of the other. They imagine a "perception-action mechanism" that is one of the basic features of neural organisation, and which provides the necessary "hardware" for the evolution of empathy. Thus behavioural matching is seen as a key to all phenomena that rely on state-matching or social facilitation, including empathy. It also follows that mammals, and more specifically group-living mammals should be able to show the basic features of emphatic behaviour.

4. EMPATHIC BEHAVIOR IN ANIMALS

Although, animals have been often credited with some capacity for empathy in some scientific circles these ideas have not found their place in main stream research, and very often any claim for empathic behaviour was dismissed as being anthropomorphic.

Recent work on mice indicates, however, that animal models of empathy might have some general validity. After having observed object mice that received electric shock paired with a tone stimulus, subject mice displayed various forms of distress to the same tone and also to the tone-shock presentation [1]. This suggests that the behavioural (including vocalisation and odours) cues displayed by the objects were powerful stimuli in evoking similar inner state in the subjects. The same study also provided some evidence that the observed tendency to show empathic behaviour was associated with the general social attitudes of the mice. Mice from a strain with more social affiliative tendencies displayed also more empathic behaviour. In another experiment it was demonstrated that observing object mice in pain intensifies the response of subjects to pain [4].

Similar studies were also run with rhesus monkeys. Subject monkeys learnt to stop shocking object monkeys by pressing a bar, and this behaviour could be also evoked by showing pictures of shocked monkeys [6]. Subject monkeys also withhold pulling chain for food if this also resulted in object monkeys being shocked [12].

Not surprisingly chimpanzees are in the focus of many studies on empathy. They also react empathically to pictures or videos showing con-specifics who display emotional behaviour (e.g. 7). Importantly, they also react to objects (e.g. needles used to injection) and to positive emotions when presented on pictures. However, in the case of the former the role of direct experience with needles cannot be excluded. In contrast to other animals studies so far only chimpanzees were found to respond also empathically to "positive" stimuli (e.g. play), that is, they displayed matching emotions.

Reading emotional expression of a group mate could also provide more direct information about the environment. In a social learning situation infant monkeys will avoid novel objects if they observe that the mother is looking fearfully at these objects [5]. In similar lines younger monkey can also learn the novel objects are not dangerous. In a reverse case infant monkeys encountering a novel object might look at the face of their mother. The phenomenon described as "social reference" provides some evidence that the emotion displayed by the adult influences the future behaviour of the infant toward the object. Both types of interactions play a major role in learning about the environment in human infants.

5. THE EMPATHIC CIRCLE

Research on empathy differentiates the "object" and the "subject". Empathy is attributed to the subject if it matches its inner state to that of the object. However, this view is too simplistic for many reasons.

Both Hoffman's [3] and Preston and de Waal's [8] definition of empathy is problematic because they refer to the *"the attended perception of the object's state".* Importantly, the subject has no means to perceive the object's "state". It can only observe the behavioural cues which are associated with the actual inner state of the object, and can only infer the underlying inner state. This distinction is important because the aforementioned authors envision a deterministic relationship between the inner state and the behavioural cues. In reality however the relationship is more complex. First, there is no evidence that inner states are matched directly on a set of behavioural cues. Some inner states may be never revealed at the behavioural level. Second, behavioural cues are probably constrained in revealing exactly any inner state, and thirdly, "information" is also lost in the perceptual process. Thus the subject can only infer, judge, or approximate the inner state of the object through attending behavioural cues (visual, acoustic, chemical etc.).

Importantly, the "object-subject" view is based on a third person perspective, and empathy is visualised as a uni-directional process. However, based on the above definition it is very difficult to discriminate "empathy" from "communication". Communication is also defined as having a sender, which by the means of specific behavioural cues, influences the behaviour of the receiver. This is especially problematic if we find that showing pictures of playing object animals releases playful behaviour from the subjects. What are the distinctive features of this interaction that differentiate communication from empathic behaviour?

A further problem is that it is not clear how the previous experience of the subject influences empathic behaviour. For example, seeing a needle could also release fear because own experience with a pain. In many experiments it is also not clear that the subject is exposed only to the actual emotional behaviour cues or they also witness how the object actually arrived at a given emotional state.

Finally, models of empathy reflect only rarely on the problem whether the object recognises the empathic behaviour of the subject. If empathy has an important role in inter-subjective relationships then there is a need of mutual recognition of empathic behaviour. This also follows from describing empathy as a form of altruism. One would expect that the behaviour of the subject gains a further advantage (also from an evolutionary point of view) if the object can recognise the empathic component. Only in this case can one assume that empathy provides a foundation for inter-subjective relationships.

6. CATEGORIES AND FUNCTIONALITY OF EMPATHY

Preston and de Waal [8] distinguished 6 levels of empathic behaviour (emotional contagion, sympathy, empathy, cognitive empathy, prosocial behaviour. They used three aspects to differentiate among these levels. They asked whether the empathic behaviour reflects a matching of the inner state, whether the subject actively acts on the object (e.g. "helping"), and whether there is some evidence for self-other distinction. As indicated above this and similar types of categorisations put an emphasis on the inner state matching and thus fail to distinguish some simpler forms of empathy from communicative interactions. Consider the case for the empathy of pain in mice cited above. One could assume that behaviour associated with pain functions in the same way as alarm signals. Alarm signals are produced by animals that witness some danger in their environment. They not only affect the behaviour of the other members in the group but also change their inner state. Visual, auditory, olfactory cues associated with pain could also have a similar effect on the subject. Interestingly, there are such alarm systems in fish. The attacked and physically harmed individual releases pheromones which initiate flight reactions from the group members.

In our view empathic behaviour can be separated from communication if we include that the subjects should pay some cost for being empathic. Thus "mirroring" or "matching" behaviour or "inner states" does not seem to fulfil criteria for empathic behaviour. By "cost" we mean that the actual matching of behaviour (or change in the behaviour) and/or inner state may not be in the own interest of the subject or, in reverse, it can be shown that by being empathic the subject investments in a personal relationship. Such cases usually involve interaction are often referred to as "consolation", "helping".

7. AFFECTIVE COMPUTING AND EMPATHIC AGENST

Research on information technology has explored for long time how emotional interaction may facilitate human-computer or human-robot (=human-machine) interaction. This lead to the emergence of a field called "affective computing" which draws it theories from the psychology of human of emotion and communication. Given the fact that the scientific understanding of human emotions is quite limited, affective computing has a very ambitious research goal when it tries to explore the possibilities of mutual communication between humans and machines based on stimuli and behavioural cues that have emotional valence. "Empathic agents" are often defined as artificial systems that are able to engage in mutual empathic communication of humans. Today it seems that there are both theoretical/conceptual and practical problems in achieving this goal. Space does not permit to reflect on all issues, however, a few important aspects derived from the above discussion on evolutionary are listed.

For many researchers empathy equals mirroring of inner states. Importantly, this is not the case for human-machine interaction because the inner state of the artificial agents and humans do not match. The common origin of species (at least for the case of mammals) provided an important argument for common processes that underlie animal and human emotions and empathic behaviour. Thus the artificial system must rely on the ability to mimic both emotional states and empathy by displaying behavioural cues for the communicative interaction. (Since this discussion is based on evolutionary comparison no attempt is many to include the utilisation of linguistic interaction in empathic interactions.). Importantly however both the design of these communicative behaviours and the recognition of the human equivalents are problematic technically.

Affective computing relies on models of human emotions. The trend is also to utilise human-like behavioural cues for the interaction which might actually decrease believability, especially in the case of robots.

The evolutionary model of empathy is based on a similarity relationship between the object and the subject at the ecological level and on familiarity at the individual level. According to Preston and de Waal [8] the bodily similarity between the interacting partners and the perceived familiarity to the object individual increases the tendency for empathic behaviour in subjects. Both arguments seem to make difficult the design of human-machine empathic interaction.

Previous discussion also indicated that empathy is more than just noticing emotions of the other and reflecting on them. For example, the subject has to have some means to infer that the object is in the position to have similar past experience. In this case the situation is very different in the case of virtual agents are robots. Humans may attribute ("fantasy") very similar capacities to a virtual agent, which looks and behaves very similar to them. In this case they do not only perceive the bodily similarity but also potential similarity in personal experience by mental attribution. However, even this may not be enough because virtual agents are probably never being confused with "real" agents. In the case of robots the believability is very questionable because the discrepancy between bodily similarities between objects and subjects, the mutual recognition (and display) of emotions, and the understanding that present robots are not in the position to have similar past experiences as their human partner.

8. ACKNOWLEDGMENTS

This research is supported by the EU FP7-ICT-2007 project (LIREC: 105554).

9. REFERENCES

[1] Chen, Q., Panksepp, J. B., Lahvis, G.P. (2009). Empathy is moderated by genetic background in mice. PLoS, 4. 1-14.

[2] Darwin C. (1999/1872). The expression of the emotions in man and animals. FontanaPress. London

[3] Hoffman, M. L. (2000). Empathy and moral development: Implications for caring and justice. Cambridge University Press.

[4] Langford DJ, Crager SE, Shehzad Z, Smith SB, Sotocinal SG, et al. (2006). Social modulation of pain as evidence for empathy in mice. Science 312, 1967–1970.

[5] Mineka, S., Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology: General* 122, 23-38.

[6] Mirsky, I. A., Miller, R. E., Murphy, J. V. (1958) The communication of affect in rhesus monkeys. *Journal of the American Psychoanalytic Association* 6, 433-441.

[7] Parr, L.A. (2001). Cognitive and physiological markers of emotional awareness in chimpanzees. *Animal Cognition*, 4, 223-229.

[8] Preston, S.D., de Waal F. B.M. (2002). Empathy: its ultimate and proximate bases. *Behavioural Brain Sciences*, 25, 1–72.

[9] Rice, G. E. J., Gainer, P. (1962). "Altruism" in the albino rat. *Journal of Comparative & Physiological Psychology* 55, 123-125.

[10] Tinbergen N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433.

[11] Trivers RL. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57

[12] Wechkin, S., Masserman, J. H., Terris, W., Jr. (1964). Shock to a conspecific as an aversive stimulus. *Psychonomic Science* 1, 47-48.