Attachment Theory and Artificial Cognitive Systems[1]

Dean Petters
Aston University
d.d.petters@aston.ac.uk

John Bowlby formulated the core of Attachment Theory whilst working in a multidisciplinary team of childcare professionals, psychoanalysts, ethologists and other researchers at the Tavistock Clinic for over thirty years after the Second World War (Holmes 1993). From its inception Attachment Theory integrated concepts from academic fields as diverse as Ethology, Cognitive Science and Psychoanalysis. Attachment Theory attempts to set out descriptions of a set of observable behaviours related to social and emotional attachment in animals and humans and the cognitive mechanisms that give rise to these observable phenomena. It can therefore be presented as incorporating theoretical components that can be termed the behavioural and cognitive components of Attachment Theory. The behavioural component of Attachment Theory is valuable for researchers interested in engineering artificial attachment systems because the set of attachment behaviours that have been observed and recorded over the last half century provide numerous behavioural scenarios. These can act as standardised specifications of requirements for the engineering process allowing artificial attachment systems to be focused in their design and evaluation (Petters 2004, 2006). The cognitive component of Attachment Theory provides a point of departure for contemporary engineering theories. Designers of artificial attachment systems can use the information processing explanations for attachment phenomena that already exist within Attachment Theory to constrain their simulations. In this case the artificial systems are using simulations to explain natural systems. Or designers can reproduce attachment behaviours with novel and less biologically plausible information processing structures and mechanisms. In this case artificial attachment systems may be being produced primarily in pursuit of engineering goals.

Section 1 -The behavioural component of Attachment Theory

The offspring of many animals, including humans, show a tendency to gain and retain physical proximity to their main carer, usually their mothers. This is not surprising, as for many of these species the mother feeds her offspring, and infant animals have to approach to be fed. Some additional proximity seeking by the offspring to its mother may be due to secondary reinforcement effects that are driven by the primary reinforcement of gaining food. However, animal studies have shown proximity seeking to a main carer can be unrelated to any reinforcing effect of feeding. Lorenz's research on imprinting in Geese and Harlow's research with infant monkeys and wire-frame mother substitutes are amongst a number of studies that show animals can imprint or attach to individuals or objects that do not provide a source of food (Lorenz 1952, Harlow 1958, Bowlby 1969|1982).

From the late 1930's Bowlby was undertaking human studies on attachment that can be seen as a parallel to Lorenz and Harlow's research on attachment processes in animals. Bowlby was interested in issues such as: the separation distress exhibited by children when they or their mothers were absent due to the infant or mothers hospitalization (Bowlby and Robertson 1952); the effect of early maternal deprivation on later development (Bowlby 1944); and the effect of grief and mourning for the bereavement of a carer that occurs in an individuals infancy (Bowlby 1960). Bowlby's description of attachment behaviours and the theoretical framework he developed to explain this body of behavioural phenomena were detailed in his Attachment Trilogy (Bowlby 1969|1982, 1973, 1980). The key insight in interpreting the attachment behaviours that Bowlby described is to see the attached infant or child as using their carer as a secure base from which to

1 Much of the material in this briefing is adapted from Petters (2006).

explore.

When infants develop the ability to crawl, and later to walk, they can explore the world more effectively and the rate at which they acquire knowledge about the broader world accelerates. However, this new found ability to explore brings with it the potential for access to many more hazards. Anderson (1972) undertook a naturalistic study of toddlers between the ages of 1 and 2 which investigated how infants balance the opportunity for exploration and the security provided by their carers. In London parks, infant and carer pairs were observed without their awareness, in observation periods averaging 15 minutes' duration. The study attempted to focus only on the exploration/security balance, therefore only cases where the pair were far enough away from interruption by other people were included in the study. Also episodes where the infant ceased exploration due to the carer attracting the infant with toys or food were excluded. This study found that most infants kept within a line of sight of their carers, and used their carers as secure bases from which to explore. It was usually at the infant's rather than carer's instigation that the infant returned to `base'.

In contrast to Anderson's study of naturalistic infant exploration, Ainsworth *et al*'s (1978) study describes the development of the Strange Situation Experiment. The Strange Situation Experiment is not an experiment where subjects are randomly assigned to different conditions in the laboratory. Rather, it is a standardised laboratory procedure that presents all infants with the same controlled and replicable set of experiences. Different outcomes arise from infants possessing different predispositions for behaviours related to attachment. These predispositions have been primarily gained from the infant's particular experiences with their carer over the previous year. The laboratory procedure consists of eight episodes of three minute duration which are designed to have the effect of activating and intensifying the attachment behaviour of one-year-old infants in a moderate and controlled manner. Some episodes involve the infant meeting an unfamiliar `stranger' in the laboratory, other episodes involve the mother being removed from the room. In all of these episodes the infant's behaviour is carefully recorded from behind a two-way mirror. In the final episode of the Strange Situation Experiment the mother returns to her one-year-old infant after the infant has been left alone for three minutes in the unfamiliar setting.

After this final reunion some infants will show some initial distress but approach the carer in a positive manner, seeking and receiving physical contact, and then quickly return to play and exploration. Others may show little distress, will not seek contact and may avoid their carer's gaze and physical contact, instead resuming play and exploration but with less concentration. Still others may respond by being overly passive or showing anger towards the carer and not resuming play and exploration.

Ainsworth *et al* (1978) found that when the measures of infant response in the Strange Situation were evaluated the infants were found clustered into three major categories of attachment style, labelled: Avoidant (type A), Secure (type B) and Ambivalent (type C) (More recent studies have categorised a fourth type of disorganised attachment style that forms a very small proportion of infants in the general population). Ainsworth *et al* (1978) found that an infant's response to reunion in the Strange Situation can act as a shorthand for the infant's home relationship with their carer. This is because the pattern of responses made by infants in this particular episode of the Strange Situation correlate most strongly with patterns of maternal behaviour and infant responses intensively observed throughout the previous year.

Avoidant infants respond to their carer on reunion in the Strange Situation by not seeking contact or avoiding their carer's gaze or avoiding physical contact with her. These children return quickly to play and exploration but do so with less concentration than secure children. Whilst playing they stay close to and keep an eye on their carer. It may seem that they are not distressed or anxious in

the Strange Situation. However, studies which measured physiological stress correlates of infants in the Strange Situation demonstrated stress levels of avoidant infants were at least as high as the secure and ambivalent groups (Hertsgaard *et al* 1995). In comparison with average levels across all groups: A type carers were observed at home being consistently less sensitive and providing more physical contact of an unpleasant nature; at home these infants were more angry and they cried more. However, in the reunion episodes of the Strange Situation these infants showed the least anger and crying.

Secure infants respond to their mothers on reunion in the Strange Situation by approaching them in a positive manner. They then return to play and exploration in the room quickly. They received care at home which can be summarised as being consistently sensitive. In comparison with average levels across all groups: B type carers were observed at home being more emotionally expressive and provided less contact of an unpleasant nature; at home these infantswere less angry and they cried less.

Ambivalent infants respond to their carers on reunion in the Strange Situation by: not being comforted and being overly passive or showing anger towards their carers. These children do not return quickly to exploration and play. They received care at home which can be summarised as being less sensitive and particularly inconsistent. In comparison with average levels across all groups: C type carers were observed at home being more emotionally expressive; they provided physical contact which was unpleasant at a level intermediate between A and B carers and left infants crying for longer durations; at home these infant's were more angry, and they cried more.

Since the development of the Strange Situation as a method to investigate individual differences in attachment style, various theories have been put forward to explain its correlational results that link early carer and infant behaviour with later infant behaviour (Goldberg 2000, chapters 4 and 5). The two main theoretical camps are: firstly, those which emphasise evidence that the correlations are due to temperamental traits possessed by the infant that are considered to be innate; and secondly, those that emphasise evidence that infants are adapting to the caregiving environments provided by their mothers. A recent review of several meta-analyses that covered hundreds of studies on thousands of participants found that maternal sensitivity is a more important factor in the formation of infant attachment than infant temperament (van Ijzendoorn and Bakermans-Kranenburg 2004.)

The Strange Situation Experiment allows individual differences in attachment status to be measured in infancy. Since the development of this procedure a number of other procedures have been developed to classify individual differences in attachment behaviour later in childhood (Soloman and George 1999) and in adulthood (Hesse 1999). The behavioural component of Attachment Theory thus constitutes a rich and varied source of behavioural data that can act as scenarios for designers of artificial attachment systems. The cognitive component of Attachment Theory provides mechanistic explanations for many of these observed patterns of behaviour.

Section 2 - The Cognitive Component of Attachment Theory

The key construct that Attachment Theory uses to explain how and when actions are selected is the concept of a control architecture formed by interacting behaviour systems. Behaviour systems as described by Bowlby possess a number of properties that can constrain simulations of attachment behaviour (Bowlby 1969|1982 , Cassidy 1999 , Petters 2006).

Each behaviour system carries out a species specific function

Bowlby suggested that infants possess a *"species-specific system of behaviours that leads to certain predictable outcomes which contribute to reproductive fitness."* (page 5 Cassidy 1999). Behaviour

systems control a broad set of behaviours. In animal ethology, behaviour systems are theorised as controlling behaviours such as mating, fighting and feeding. Each behaviour system carries out a species specific function, and has been selected for this function in the evolutionary past. Hinde recounts that:

> ``*The concept of a behavioural system is, in fact related to one meaning of the term instinct. [...] It has been used in a rather special sense by ethologists to refer to systems postulated as controlling a group of behaviour patterns that together serve to achieve a given biological end" (Hinde 1983, page 57).*

The behaviour systems that Bowlby linked to attachment behaviour in human infants are the attachment, fear, sociability and exploration systems (Bowlby 1969|1982). In the view of Attachment Theory, behaviours resulting from the attachment behaviour system and the fear system have the predictable outcome of increasing the proximity of the attachment figure to the infant, the principal means by which infants gain security. The exploratory behaviour system activates behaviours that result in learning and the sociable system results in social interaction. The result of this theoretical commitment is that the agent based simulations reported in this thesis possess goal or motive based architectures. When a behaviour system in Bowlby's sense becomes activated it has been interpreted as requiring the activation of a goal or motive, though these may be implicitly represented.

The behaviour systems most closely related to attachment are inherently motivated.

Infants will work to experience exploration, socialisation and security because these outcomes can be considered primary drives. They are not activated as the by-product of any more fundamental process (Cassidy 1999). This has come to be the mainstream view in child development, but this was not always the case. For example Smart and Smart (1967) held the view that infants had to have returning behaviours reinforced by their carers or else *"there would be no more babies, since babies creep and toddle right into danger"*.

This inherent motivation to explore was recognised by Piaget, who described how a child's interest in exploration drives cognitive development (Marvin and Britner 1999). What this means is that infants don't learn, by some process of association or re-inforcement, to use their carers as secure bases from which to explore the world (Bowlby 1969|1982). Running away, freezing and using a carer as a secure base are all behaviours that humans and other animals may instinctively `know' to do when faced with danger of particular types. This has important implications for the designers of artficial attachment systems.

Behaviour systems possess a flexible repertoire of behaviours.

What defines a behaviour system is not a set repertoire of behaviours but the outcomes that predictably follow from the carrying out of those behaviours. Similar behaviours may be produced by different behaviour systems. For example, behaviours such as locomotion can serve more than one system, such as the attachment and exploratory system. Also any given behaviour system may produce a wide range of differing behaviours. In the attachment system, if the infant possesses the goal of increasing its proximity to a carer the infant may cry to bring the carer closer or crawl towards the carer themselves. This is an example of behaviours within systems being interchangeable with other functionally equivalent behaviours.

According to Attachment Theory, the attachment behaviours used to gain proximity range from:

subtle signals, such as gazing towards a carer, to overt signals, such as calling a carer, and active behaviours, such as locomotion. Exploratory behaviours range from locomotion to object manipulation and have the predictable outcome of improving the infant's ability to manipulate the external world. The consequence of this theoretical commitment is that is has allowed the simulations to be set at a high, goal/motive oriented level of description.

Behaviour systems in humans involve a hierarchy of forms of information processing

In his description of Behaviour Systems Bowlby invoked a variety of mechanisms. These ranged from simple mechanisms, such as reflexes and fixed action patterns, through goal corrected mechanisms, to complex mechanisms and forms of representation, such as Internal Working Models and natural language. Information processing structures and mechanisms that Bowlby (1969|1982) } posited in Attachment Theory as putative explanatory mechanisms for attachment behaviours include:

- Ethological concepts and mechanisms, such as Behaviour Systems, Reflex Actions and Fixed Action Patterns which can interact in complex ways by chaining and alternation;

- Concepts from the study of Control Systems such as feedback and goal directed mechanisms;

- Concepts from AI and Cognitive Science such as Internal Working Models (IWM's) and hierarchical organisation and control of behaviour using complex representational forms such as natural language.

Reflexes are behaviours with a highly stereotyped form. Once activated by a stimulus at a specific threshold they are ballistically carried to completion (Marvin and Britner 1999). Fixed action patterns are similar to reflexes because they are stereotyped but differ from reflexes because they are open to learning. The thresholds for activation and termination adapt according to the state of the organism and past experiences. Fixed action patterns are also less ballistic, with for example, proprioceptive feedback during execution. Examples of this type of behaviour include grasping, crying and smiling (Marvin and Britner 1999). Reflexes predominate in the first few months after birth and fixed action patterns predominate from three months until the middle and end of the first year. The reflexes and fixed action patterns that infants perform may seem a very simple form of control but are highly effective in eliciting adult actions that benefit the infant. Different reflexes and fixed action patterns are coordinated together, the sum of these behaviours working together is greater than their parts. As Marvin and Britner (1999) note:

> ``*The immediate effect of any behaviour is to bring about a change in the environment, which itself serves as an activating condition for another behaviour, often forming a lengthy sequence with an eventual outcome that is necessary for the individual's survival." (Marvin and Britner 1999, page 48).*

These complex patterns produced by fixed action patterns can be mistaken for behaviours directed by goal corrected mechanisms because of the sensitive matching of response to stimuli. The goal corrected phase of attachment commences from the middle to the end of the first year. When a mechanism is goal-corrected it is updated or retaken according to feedback on how well the goal has been satisfied.

The concept of Working Models came to Attachment Theory from the work of Craik (1943). In a

broader sense Working Models are not confined to attachment but apply to all representative models of the world. In his work Bowlby restricted the term Internal Working Models (IWM's) to models of self and other in attachment relationships. IWM's capture the relation-structure of phenomena, not every aspect of reality but enough to make possible the evaluation of alternative actions. These include spatio-temporal causal relations among the events, actions, objects, goals and concepts represented. IWM's of attachment are what hold an infant's expectations of the levels of predicted availability and responsiveness for a given carer. These expectations are derived from the carer's past performance. IWM's of self and attachment figure develop in a complementary manner. For example if the carer is responsive the self is valued.

Bowlby's invocation of such wide range of forms of representation and types of mechanism has meant that the design considerations in these areas have been minimally constrained by commitments from Attachment Theory.

Behaviour systems are described at a high level of implementation.

Behaviour systems are described at the level of software or virtual machine operation, as opposed to a neural level of description. Hinde (1983, page 57) explains:

> ``*Although the behavioural systems were postulated to explain the observed behaviour, there was no necessary implication of isomorphous mechanisms in the brain. The explanation is a software one, comparable to a computer program which performs a particular job irrespective of the details of the hardware of the computer into which it is fed.''*

Examples of artificial attachment in software or robotic simulations

Bischof (1975) presented an early example of a simulation of attachment behaviour with a software simulation of infant approach and avoidance behaviours. This work didn't closely follow Bowlby's ideas on a control system formed of independent behaviours. Its control mechanisms were instead based upon more abstract cybernetic control circuits. More recently, several research groups working with robots have created control systems that re-create patterns of attachment behaviour, but again do not do so using control architectures based upon Bowlby's Behaviour System architecture. Likhachev and Arkin (2000), and Arkin (2005) describe experiments conducted with autonomous software agents and robots that are linked to Attachment Theory by using this theory as a source of organising ideas and motivation. The goal of this work was to design artefacts that would exploit the use of 'comfort zones' in the exploration of their surroundings. As Arkin notes:

> *"The intent here is not to create a robotic model of a human child, but rather to produce useful behavior in autonomous robotic systems. While robots are not children, they are nonetheless advantages to maintaining an attachment bond with certain individuals or objects, (e.g. Caregivers, owners, military base, fuel supply, or familiar end users), as they typically satisfy the robit's endogenous needs (e.g. energy) whilst also providing a high-high level of familiarity and predictability in the environment."* (page 16, Arkin 2006).

An engineering approach to the creation of imprinting and other affective and social recognition competences in robots is described by Canamero, Blanchard and Nadel (2006); Hiolle, Canamero and Blanchard (2007); and Hiolle and Canamero (2007). This work concentrates on the requirement of these robots reproducing external patterns of secure-base behaviour in real physical environments without attempting to reproduce the kinds of internal hierarchical control structures

that Bowlby postulated.

Petters (2006) reports the implementation of a number of simulations that reproduce patterns of attachment behaviour described in scenarios taken from Anderson's (1972) and Ainsworth et al's (1978) empirical studies. The architectures presented by Petters (2006) fall into two broad categories. The simpler architectures possess action selection mechanisms somewhat similar to Cooper and Shallice's (2001) model of Contention Scheduling. In these architectures the specified patterns of behaviour arise as a result of reactive goals competing against each other for control of motor actions. More complex architectures have deliberative subsystems built upon the reactive action selection mechanisms. These deliberative subsystems can create several behavioural strategies in the form of simple linear plans that are chains of states of the world linked with actions. These plans can then be evaluated, selected and executed. Actions arising from reactive subsystems can be thought of as pre-potent responses. Execution of deliberated actions involves inhibition of these reactive actions.

Existing examples of artificial systems form a limited set of relatively simple architectures. Future work will involve creation of more sophisticated architectures, which might incorporate greater detail in attentional systems, learning mechanisms, or in the diversity of representations and integration between high and low level mechanisms (Dewitte *et al* 2007).

References

Ainsworth, M., Blehar, M., Waters, E., and Wall, S. (1978) *Patterns of attachment: A psychological study of the strange situation.* Erlbaum, Hillsdale, NJ.

Anderson, J.W. (1972). Attachment behaviour out of doors. In, N. Blurton-Jones, (eds) Ethological studies of child behaviour, pages 199-215. Cambridge University Press, Cambridge, UK.

Arkin, R.C., (2005) Moving Up for Food Chain: Motivation and Emotion in Behavior Base Robots. In, J.M. Fellous and M.A. Arbib (Eds) *Who Needs Emotions? The brain meets the Robot.* Oxford University Press

Bischof, N., (1975) A Systems Approach toward the Functional Connections of Attachment and Fear, *Child Development* 46, 801-817

Bowlby J. (1944). Forty-four Juvenile Thieves: Their Character and Home Life. International Journal of Psychoanalysis, 25: 1-57 and 207-228; republished as a monograph by Bailliere, Tindall and Cox, London

Bowlby J. And Robertson, J., (1952). A two year old goes to hospital *Proceedings of the Royal Society of Medicine*, 46: 425-7.

Bowlby J. (1960). Grief and Mourning in Infancy and Early Childhood. *The psychoanalytic study of the child XV*: 9-52.

Bowlby J. (1969|1982). *Attachment and Loss*, volume 1, *Attachment*, Basic Books, New York.

Bowlby J. (1973). *Attachment and Loss*, volume 2, *Separation: Anxiety and Anger*, Basic Books, New York.

Bowlby J. (1980). *Attachment and Loss*, volume 3, *Loss: Sadness and Depression*, Basic Books, New York.

Canamero, L., Blanchard, A., and Nadel, J., (2006). Attachment Bonds for Human-Like Robots. International Journal of Humanoid Robotics, 3(3): 301-320

Cassidy, J. (1999). The nature of the child's ties. In *Handbook of Attachment*, eds J. Cassidy and P.R. Shaver, pages 3-20. Guilford Press, London.

Cooper R. and Shallice T. (2001). Contention Scheduling and the control of routine activities. *Cognitive Neuropsychology*, 17(4):297-338

Craik K. (1943). *The Nature of Explanation*, Cambridge University Press, London-New York.

Dewitte M, De Houwer J, Koster EH, Buysse A. (2007) What's in a name? Attachment-related attentional bias. *Emotion*. 7(3):535-45.

Goldberg, S. (2000). *Attachment and development*, Arnold, London.

Harlow, H. (1958). The nature of love. *American Psychologist*, 13:673-85.

Hertsgaard, Gunnar, M., Erickson, M.F. and Nachmias, M. (1995). Adrenocortical responses to the strange situation in infants with disorganised/disorientated attachment relationships. *Child Development*, 66:1100-06

Hesse E. (1999) The Adult Attachment Interview: Historical and Current Perspectives. In *Handbook of Attachment*, eds J. Cassidy and P.R. Shaver, pages 395-434. Guilford Press, London.

Hinde, R. (1983). Ethology and Child Development. In, *Handbook of Child Psychology*. Ed *P.H. Mussen*, pages 27-93. J. Wiley and Sons, New York.

Hiolle, A., Canamero, L., and Blanchard, A. (2007). Learning to Interact with the Caretaker: A Developmental Approach. In A. Paiva, R. Prada, and R.W. Picard (Eds.), Proc. Second International Conference on Affective Computing and Intelligent Interaction (ACII 2007), LNCS 4738, pp. 425-436. Berlin Heidelberg: Springer-Verlag.

Hiolle, A. and Canamero, L. (2007). Developing Sensorimotor Associations Through Attachment Bonds. In C. Prince, C. Balkenius, L. Berthouze, H. Kozima, M. Littman (Eds.), Proc. 7th International Conference on Epigenetic Robotics (EpiRob 2007), Lund University Cognitive Studies, 134, pp. 45-52.

Holmes J. (1993). *John Bowlby and Attachment Theory*. Routledge, London.

Likhachev, L. and Arkin, R.C., (2000), Robotic Comofrt Zones. In, *Proceedings of SPIE: Sensor Fusion and Decentralized Control in Robotic Systems*, 27—41.

Lorenz K.(1952), *King Solomon's Ring.* Methuen, London

Marvin, R.S. and Britner, P.A. (1999). Normative development: The ontogeny of attachment. In *Handbook of Attachment*, eds J. Cassidy and P.R. Shaver, pages 44-67. Guilford Press, London.

Petters, D (2004) Simulating Infant Carer Relationship Dynamics. In *Proceedings of the AAAI Spring Symposium, (March 2004, Stanford): Architectures for Modelling Emotion - Cross-Disciplinary Foundations*, number SS-04-02 in AAAI Technical reports, pages 114-122, Menlo Park, CA: AAAI Press. http://www.cs.bham.ac.uk/~ddp/simbaby.pdf

Petters, D. (2006), Desiging Agents to Understand Infants. *Ph.D. thesis in Cognitive Science*, Department of Computer Science, University of Birmingham. http://www.cs.bham.ac.uk/~ddp/petters-phd-2006.pdf

Smart, M.S., and Smart, R.C. (1967). *Children: Development and Relationships*. Macmillan, New York

Soloman J. And George C. (1999) The measurement of attachment security in infancy and childhood. In *Handbook of Attachment*, eds J. Cassidy and P.R. Shaver, pages 3-20. Guilford Press, London.

Van Ijzendoorn, M.H. and Bakermans-Kranenburg, P.M. (2004). Maternal Sensitivity and infant temperament. In Theories of Infant Development, eds G. Bremner and A. Slater, pages 233-257. Blackwell Press Oxford.