The attachment control system and computational modeling: Origins and prospects

PETTERS, Dean

Published version

PETTERS, Dean (2019). The attachment control system and computational modeling: Origins and prospects. Developmental Psychology, 55 (2), 227-239.

The Attachment Control System and Computational Modeling: Origins and Prospects

Dean David Petters

University of Wolverhampton

Author Note

Dean David Petters, Department of Psychology, University of Wolverhampton, Wulfruna Street. Wolverhampton, UK.

Correspondence: Dean David Petters, Psychology, Institute of Sport and Human Sciences, Millennium City Building, Wolverhampton WV1 1LY; Email: D.D.Petters@cs.bham.ac.uk

ABSTRACT

From his first attempts to explain attachment phenomena in the 1940s through his *Attachment and Loss* trilogy (trilogy; 1969|1982, 1973, 1980), John Bowlby reformulated the theoretical underpinnings of attachment theory several times. He initially attempted to explain attachment phenomena in psychoanalytic terms. Then he invoked ethological theory in the explanation of how and why people behave as they do in close personal relationships. The mature theoretical framework that he presented between 1969 and 1982 in the attachment and loss trilogy retained strengths and insights, ultimately situating them within an overarching control systems framework. This paper describes key stages in Bowlby's theoretical development, with particular emphasis placed on the emergence of control systems theory as a cornerstone of the mature theory. It also compares Bowlby's control systems approach to contemporary cognitive science approaches. It concludes by suggesting how Bowlby's control systems formulation could evolve along the path opened up by contemporary work in computational modeling and how it could benefit by doing so.

Keywords: attachment theory; Computational Modeling; Cognitive Science; Attachment control system; Internal Working Model; Cognitive architecture

THE ATTACHMENT CONTROL SYSTEM AND COMPUTATIONAL MODELING

This paper is an analysis of how John Bowlby developed the 'attachment control system' concept. Bowlby's use of the concept originated as an attempt to explain a complex cluster of behavioral and psychological phenomena related to attachment relationships, without invoking explanations in terms of psychoanalytic constructs. Bowlby's approach assimilated an armory of modeling techniques from cybernetics, artificial intelligence and other scientific approaches to explaining these phenomena. In doing so, he employed both emerging cognitive theories and introduced concepts and explanations which were ahead of his time. Indeed, many of his explanations have a very contemporary cognitive science 'feel' (Petters, 2016). However, he faced one critical obstacle in presenting cognitive science complexity as a substitute for psychoanalytic complexity - namely, that cognitive science was in its infancy. At the time, few models spoke directly to the kinds of psychoanalytic processes he wanted to re-describe in cognitive terms.

Bowlby described his foray from psychoanalytic theory into ethology, and then into a variety of research areas concerned with human information processing, as an attempt to provide a scientifically respectable foundation for attachment theory (Bowlby 1981). A conceptual analysis of the development of attachment theory provides useful insights into Bowlby's strategy, and into the convergence of his ideas with the undertaking we recognize today as cognitive science.

Bowlby (1969, p. 43) viewed his control systems framework as an initial explanatory model to be extended and deepened, and ultimately integrated with scientific explanations at other levels of abstraction. Recent neuroimaging results presented by Bretherton and

Munholland (2016) support this approach. However, since the information processing models with which Bowlby illustrates his attachment explanations are relatively simple, it can seem that he was presenting these ideas as a metaphor that could be invoked to differentiate attachment theory from psychoanalysis, and then be safely ignored (Bretherton 1985, p. 12-13; Bretherton 1999, p. 1; Bretherton and Munholland 2016, p. 63 and p. 68; Hinde 1991, p. 378 and 397). A key contribution of this paper is to show that the control system concepts Bowlby employed can be more than mere metaphors. Using contemporary cognitive science modeling, it is now possible to implement the kind of empirically accessible, information based attachment theory that Bowlby anticipated but found just beyond his reach.

STAGES IN THE DEVELOPMENT OF ATTACHMENT THEORY

Bowlby was spurred to develop attachment theory by observing the effects separation and loss on human relationships. His data included observations of childhood evacuations during war-time (Bowlby, 1940a), the prohibition of parental hospital visits to their young children (Bowlby, 1940b), the effect of early maternal deprivation on later development (Bowlby, 1944), and the behavioral phases of grief and mourning observed in the course of long term separations during infancy (Bowlby, 1960a).

Uncomfortable with psychoanalysts use of retrospection and inference, Bowlby's initial thinking leaned heavily on such observation and description (e.g., Bowlby, J. 1951; Bowlby & Robertson, 1952); theoretical development and innovation were secondary. In addition, the target for theoretical explanation was initially broader and then focussing strategically on the mother-infant relationship. During attachment theory's long theoretical development, Bowlby imported ideas from diverse disciplines to explain phenomena related to social and

emotional attachments in close relationships (Ainsworth & Bowlby, 1991; Bowlby, 1969 / 1982; Bretherton, 1992; Holmes, 1993). He adopted new concepts to substitute for psychoanalytic constructs which he wanted to transform. Surveying the range of concepts Bowlby introduced over time, we can see he was influenced by dramatic changes in the intellectual milieu between 1950 and 1980. Explanatory concepts which were newly prominent had a particularly strong influence as Bowlby formulated elements of attachment theory at those times. The development of attachment theory during this period falls into three phases: (1) Bowlby's initial thinking in the late 1940s and early 1950s had a heavy psychoanalytic influence (Bowlby, 1956); (2) introducing concepts from ethology (Bowlby, 1958, 1960d); and (3) from cybernetics and artificial intelligence (Bowlby, 1969, 1973) and ultimately integrating with cognitive psychology (Bowlby, 1980).

Before Attachment theory: A Psychoanalytic Theory

Bowlby became interested in personality development and the key role played by an individual's early caregiving environment before he trained as a psychoanalyst ((Ainsworth & Bowlby, 1991), p 333). His belief in the significance of real life events on the course of child development set him in conflict with several psychoanalysts with whom he worked at the Tavistock Clinic in the late 1940s, resulting in his forming his own research unit in 1948 (Bretherton, 1992). At an early stage of his research career he chose to focus on the effects of early separation from the mother rather than other examples of disturbed family interaction. Bowlby made this decision due to practical considerations, as he intended to work within a scientific methodology which focused on analyzing the effect of environmental conditions. This was in contrast to adopting the retrospective case study conducted by other psychoanalysts. From Bowlby's view, separation events were

particularly suitable for research as they were an event on record, whereas at that time there was no adequate reporting or documentation for other forms of disturbed family interaction (Ainsworth & Bowlby, 1991, p. 334).

Bowlby's departure from the mainstream of psychoanalysis was also due to his awareness of problems with Freud's motivational theory. For Bowlby, this approach to motivation required revision because it was rooted in a drive theory which suggested infants were primarily focused on their inner drives and drive representations, and little interested in the social or physical environment per se (Waters, Kondo-Ikemura, Posada, & Richters, 1991). This focus inwards was in part driven by the psychoanalytic retrospective case study method which Bowlby had rejected (van der Horst, 2011). Critiques from psychology and philosophy of science also made clear that the drive theory of motivation was not tenable. It was not well supported by their own evidence, which itself was problematic, and seemed inaccessible to ordinary standards of empirical analysis and falsification (Bowlby, 1981).

In 1951 Bowlby published a landmark report to the World Health Organization which contained a substantial body of observations on the mental health of children (Bowlby, 1951). Although this report was principally a survey of empirical work, it also illustrates the manner in which Bowlby was attempting to reshape psychoanalytic theory. Bretherton (Bretherton, 1992) notes that: *"it is interesting to examine the 1951 report from today's perspective. At that time Bowlby still used the terminology of traditional psychoanalysis (love object, libidinal ties, ego, and superego), but his ideas were little short of heretical."* (Bretherton, 1992, p. 50)

Bowlby's unorthodoxy is highlighted by Bretherton ((Bretherton, 1992), p. 51):

"[The mother] *is his ego and his superego. Gradually he learns these arts himself,*

and as he does, the skilled parent transfers the roles to him. This is a slow, subtle and continuous process, beginning when he first learns to walk and feed himself, and not ending completely until maturity is reached" ((Bowlby, 1951, p. 53, quoted in (Bretherton, 1992, p. 51).

According to Bretherton, this description: *"sounds more Vygotskian than Freudian"*, ((Bretherton, 1992), p. 51). Why did Bowlby attempt a transformation of psychoanalytic terms like ego and superego rather than just rejecting all of the psychoanalytic framework outright? Psychoanalysis did possess a number of key insights into early experiences and relationships which Bowlby valued and wanted to maintain in his own approach (Waters et al., 1991). These insights include that: human infants have a complex cognitive and emotional life; attachment cannot be equated with the amount of overt behaviour toward attachment figures or with the amount or duration of protest that follows separation; and early attachment relationships are prototypes of later love relationships (Waters et al., 1991). He therefore wanted to reform psychoanalytic theory not replace it wholesale. However, Bowlby, keenly aware of the sociology of science, recognized that critics would likely throw out the genuine insights along with the untenable motivation theory. He also recognized that the useful insights about early experience and relationships were logically independent of the drive theory. Central then to his work on attachment theory at this stage was to find an alternative approach to motivation. The key here was to avoid replacing one kind of magic (drives) with another (e.g., the infant intends, signals, wants, needs, loves, etc. the mother) thereby incurring what Dennett (1981) calls "intelligence loans", presumptions of intelligence that are unlikely to be accounted for ("paid back") (Richardson, Shockley, Fajen, Riley, &

Turvey, 2010). As we shall see, first in ethology, and later in cybernetics and artificial intelligence, Bowlby found approaches that could account for what he called 'the apparently purposeful' organization of observable attachment behavior without incurring such intelligence debts.

As Ainsworth and Bowlby (Ainsworth & Bowlby, 1991) recount, it was in the early1950s that:

"Bowlby [...] *had begun a search for adequate explanation of the empirical findings, having found none in current psychoanalytic theories to account for young children's responses to separation and reunion, or indeed how the tie to the mother develops. At this point Konrad Lorenz's work on imprinting became available in translation. Sensing its possible relevance to his problem and encouraged by Julian Huxley, Bowlby began delving into the ethological literature."* (Ainsworth & Bowlby, 1991, p. 337)

Though the first scientific approach that Bowlby drew upon was ethology, Bowlby's revision of psychoanalytic theory with an ethological perspective occurred gradually. The transitional nature of Bowlby's theoretical perspective at this time is illustrated by Bowlby himself in 1953, when he stated:

"I want to remark on three or four psychological processes which may be relevant, and in doing so I shall speak in a hybrid, bastard language which I have come to use, which derives from both psychoanalysis and ethology" (Bowlby, 1956, p. 183-184).

This shows Bowlby did not experience an overnight conversion from a psychoanalytic view to the scientific view provided by the then nascent disciplines of cybernetics, artificial intelligence and systems theory. However, we can see from the meetings he attended and the interactions he engaged in at these meetings that he was inexorably moved towards

viewing attachment relations in scientifically respectable information processing terms. For example, the transcript of the second meeting of the 'World Health Organization Study Group on the Psychobiological Development of the Child', in 1954, records Bowlby and Grey Walter discussing the scientific nature of the psychoanalytic superego' construct:

BOWLBY: *The superego is rather complicated and contains more than one variable*

GREY WALTER: *Can you measure the superego?*

BOWLBY: *You cannot.*

GREY WALTER: *Then what is the point of discussing it?*

BOWLBY: *I think it is useful to try and see how things relate in these psychological functions after which we are dimly trying to grope.*

(Grey Walter, 1956, p. 205).

To conclude, what Bowlby was trying to achieve by his departure from psychoanalysis was to explain attachment phenomena in a scientifically respectable manner without a wholesale rejection of the psychoanalytic perspective (Bowlby 1981). We might say that the psychoanalytic perspective acted as a kind of lens, because of the way Bowlby viewed phenomena through the filter and focus of his existing understanding, and a specification of requirements because a new scientifically respectable theory had a set of phenomena (important to psychoanalytic theory) which it must explain. In particular he wanted to explain aspects of the mother-infant relationship which were highlighted by the psychoanalytic approach. However, this was tricky because the kinds of psychological phenomena focused upon by psychoanalytic theorists are intimately related to the storehouse of theoretical mental structures and mechanisms invoked within this community.

Attachment theory Proper: An Evolutionary Ethological Formulation

After Robert Hinde (a leading ethologist) joined the seminar group Bowlby established at the Tavistock clinic in 1954, the incorporation of ethological theory in Bowlby's conceptual development deepened. This collaboration helped contribute to Bowlby's first formal statement of attachment theory in his 1958 paper: *'The Nature of the Child's Tie to his Mother'* (Bowlby, 1958). This initial formulation was built upon three key postulates (1) attachment behaviors are species-specific behavior patterns which he termed 'instinctual responses'; (2) the infant's attachment is mediated by a set of simple behaviors which are part of our primate evolutionary endowment, and (3) with experience these behaviors become integrated into a more complex behavioral system with the predictable outcome of establishing and maintaining proximity to the mother. ((Bowlby, 1958, p. 366). It is worth emphasizing that, although later theories provided more detail on how the attachment system develops, even this early theory presents attachment responses as constructed through interaction between infants and their caregiving environment. This was an important point because, for many, invoking evolution suggested that the attachment system was like a preformed blueprint waiting to be triggered or maturing without experience. This was not Bowlby's view at all.

Between 1958, and the 1969 publication of the first volume of the Attachment and Loss Trilogy, Bowlby published six papers that reflect significant advances in his thinking (Bowlby, 1960a, 1960b, 1960c, 1960d, 1961a, 1963). From these ethological/evolutionary insights, Bowlby crafted alternative explanations for several key psychoanalytic formulations. Bowlby showed how ideas from ethology could be used to explain the ontogeny of object relations and anxious, depressed and defensive responses which are triggered by infant-mother separations (Bowlby 1960d). He considers the ethology inspired

idea that attachment is a primary need which leads to the emergence of reciprocal social relationships during ontogeny. This theoretical innovation leads to the understanding that distress and anxiety are normal expectations when separations occur. While he still refers to many psychoanalytic constructs like ‘orality’, ‘repression’ and ‘symbolic substitutions’ he does speculate whether behavior previously thought to arise from these mechanisms might be better described by ethological concepts. For example, where psychoanalysts would explain one activity taking the place of another because of symbolic equivalence, Bowlby describes how this may occur at an intrasymbolic level due to ethological displacement activities. Two papers give an overview of behavioral phases related to attachment and loss, linking separation anxiety to protest; grief and mourning to despair and depression; and detachment to defensive processes (Bowlby, 1960c, 1961a). He makes the claim that these three types of response are phases of a single process, and each phase is best understood as part of the whole process. Bowlby claims that this insight was long to come to psychoanalysts such as Freud because they started with repression and ‘worked backwards’, rather than starting with loss and despair and then considering what might follow (Bowlby, 1960c). Bowlby also discussed links between models of motivation and mourning in infancy and adulthood (Bowlby, 1960a, 1960b, 1963). He was also clear in setting out the limitations of energy models of motivation. In arguing that an ethological framework is an improvement on psychoanalytic ideas he provided a great deal of detail on similarities between mourning in humans and in other animals. He went much further than the 1958 paper in providing details of ethological mechanisms like the environmental triggers that activate behavior. He also reflects more deeply on theoretical foundations for this ethological model of attachment, with greater consideration of evolutionary functions

related to phenomena such as the subjective painfulness of mourning, and why grief often involves anger and hatred.

Taken together these papers demonstrate Bowlby was creative in re-describing phenomena in ethological terms which had previously only been explained in psychoanalytic terms. It is noticeable how far he attempts to go with ethology as a substitute framework for psychoanalysis, in the depth and breadth of the phenomena he attempts to re-explain. In these six papers Bowlby is presenting, in a scientific manner, the whole individual person, in his or her environment, and presenting them as possessing an ethological motivational system arrived at from an ontogenetic and evolutionary trajectory. However, no complex internal dynamics are presented in the approach set out in these six papers. When Bowlby does speculate on details of an internal architecture he either sticks with the simplicity of the ethological approach to architectural concerns, or he reverts back to a selective adaptation of Freudian ideas. Therefore, if Bowlby had stopped the development of attachment theory at the ethological stage the resulting framework would have lacked the complexity needed to explain the broad collection of cognitive and affective lifespan phenomena explained by the full control systems approach set out in his 'Attachment and Loss' trilogy (Bowlby, 1969, 1969 / 1982, 1973, 1980). In summary, the ethological version of attachment theory was in several respects only part-way towards Bowlby's final characterization of Attachment theory.

The 'Full-Strength' Control System Formulation

Bowlby's final version of attachment theory, which was set-out across all three volumes of Bowlby's Attachment Trilogy (Bowlby, 1969, 1969 / 1982, 1973, 1980), involves a richer and deeper conceptualization for attachment phenomena. Within a control system

framework, Bowlby's three volume trilogy included information processing concepts such as homeostasis, hierarchical plans, internal working models, selective attention, and meta-cognition. These concepts provide a coherent, integrated model for explanations of a range of attachment phenomena. Van der Horst (2011) described Bowlby's adoption of an ethological framework as an *"Archimedean moment"* (van der Horst, 2011, p. 3). However, the adoption of a cybernetic and control systems framework in the late 1960s has as much claim to be a core pivotal moment as the earlier adoption of ethology in the late 1950s.

Continuity and change in theoretical developments

Bowlby reflected on the changing nature in his own theoretical approach during the 1960s, when he described the difference between his 1958 version of his theory and the version in the 1969 first volume of the Attachment and Loss Trilogy:

"The hypothesis proposed represents a development of that advanced by me in 1958. The principal change is due to better understanding of control theory and to recognition of the very sophisticated forms that behavioural systems controlling instinctive behaviour may take. In the present version of the hypothesis it is postulated that, at some stage in the development of the behavioural system responsible for attachment, proximity to mother becomes a set-goal. In the earlier version of the theory five patterns of behaviour - sucking, clinging, following, crying, and smiling - were described as contributing to attachment.

In the new version these same five patterns are still held to be of great importance, but it is postulated that between the ages of about nine and eighteen months they usually become incorporated into far more sophisticated goal-corrected systems. These systems are so organised and activated that a child tends to be maintained in proximity to his

mother. [...] *The earlier version of the theory was described as a theory of component instinctual responses. The new version can be described as a control theory of attachment behavior"* ((Bowlby, 1969 / 1982, p. 180).

Bowlby's 1969 version of attachment theory shows the continued importance of secure base behavior with an increasing role for mental representation. As a control theory, the newer framework provides a greater focus on the attachment system as directed towards outcomes as set-goals to be achieved from a flexible behavioral repertoire rather than a system that simply involves triggering preset responses. However, the new theory still includes a strong ethological influence. Although the control systems formulation was a major departure from Bowlby's early instinct theory, he retained his commitment to behavioral biology. For example, Bowlby still presents the attachment system as an instinct to form bonds and as a system that is activated by species specific patterns of care. In addition, Bowlby's new terminology of behavior systems only masks a core theoretical inheritance from his ethological instinct theory. As Hinde notes:

"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, p. 57).

Even while Bowlby was providing an alternative to psychoanalytic explanations for attachment behavior, the range and type of phenomena of interest to psychoanalysts provided much of his research agenda:

"The resulting conceptual framework is designed to accommodate all those phenomena to which Freud called attention for example, love relations, separation,

anxiety, mourning, defense, anger, guilt, depression, trauma, emotional detachment, sensitive periods in early life and so to offer an alternative of the traditional metapsychology of psychoanalysis" ((Bowlby, 1969 / 1982), p. 668).

Although Bowlby claimed that the trilogy sets out an approach which had already been fully conceived of at its initiation (Waters, personal communication), there were some limited changes in emphasis between the 1969 and 1980 publications. For example, the 1969 volume incorporated Cybernetic and AI concepts and the 1973 and 1980 volumes show a switch in emphasis to Cognitive Psychology. This change is not surprising. Bowlby noted that both his, and Freud's previous formulations of instinctive behavior, were both *"a reflection of the scientific climate of the times"* ((Bowlby, 1969 / 1982, p. 18). Clearly, to incorporate up-to-date citations in the three volumes of his trilogy as he published them would involve inclusion of some influences from contemporary theoretical sources. However, how far Bowlby was influenced merely by the nature of the ideas in current circulation at any given time should not be over emphasized. During the 30 years in which he constructed attachment theory he selected particularly suitable concepts which matched his current requirements. So, he did incorporate concepts from the current literature, but from a wide set of candidate ideas he selected those which possessed the appropriate properties and fitted within his existing framework. For example, in both the first and third volumes of the attachment trilogy psychoanalytic defensive processes are reframed in terms of how sensory inputs are processed. However, in the first volume this process is framed and referenced in terms of neurophysiological processes, in particular drawing upon the work of Magda Arnold (Bowlby, 1969 / 1982, p. 103), whereas in the third volume the same attachment phenomena are described in the newly fashionable cognitive psychological

terms of selective attention (Bowlby, 1980, chapter 4).

THE ARCHITECTURE OF BOWLBY'S ATTACHMENT CONTROL SYSTEM

In the first volume of his Attachment trilogy, Bowlby was continuing his search to consolidate the conceptual foundation for attachment theory by replacing Freud's concept of psychical energy and its discharge (Bowlby, 1969 / 1982, p. 18). He introduced the attachment control system concept for this purpose. Bowlby presented ethological mechanisms operating within the attachment control system architecture as carrying out an action selection role. According to Bowlby, what defines the attachment control system is not a set behavior repertoire but the outcomes that predictably follow from these behaviors. Similar behaviors may be produced by different behavior systems. In both the 1958 and later versions of attachment theory, the behaviors related to attachment were organized according to four behavior systems, the attachment, fear, sociability and exploration systems. Where in the 1958 theory attachment instincts were linked to the activation of particular behaviors (which can be viewed as having an implicit goal in the sense of having a predictable outcome in environments similar to the species environment of evolutionary adaptedness), in the later final theory they were linked to the achievement of particular goals which are explicitly held as representations and which can be achieved by a variety of actions.

To illustrate the internal organization of the attachment control system, Bowlby set out a range of control systems of increasing sophistication. For a simple example of a control system, which acted as a regulator by keeping a single variable constant, Bowlby presented the humble thermostat. He noted that this system was relatively static (with a pre-set goal) and unable to act as a model of even the simplest form of instinctive behavior. However, he

((Bowlby, 1969 / 1982), p. 44) showed how this control system design might be elaborated by allowing its goal setting to be determined by another control system. Such as servo mechanisms operating in a hierarchy. He also sketches out further extensions to this simple design of a greater sheer scale and complexity. In Bowlby's thinking the importance of the intimate integration of different representational forms is demonstrated when he noted:

"The mental apparatus can be thought of as made up of a very large number of complex control systems, organized in a loosely hierarchical way and with an enormous network of two-way communications between them. At the top we postulate one or more principal evaluators and controllers, closely linked to long term memory and comprising a very large number of evaluation (appraisal) scales ranged in some order of precedence. This system, or possibly federation of systems, I shall call the Principal System (s), this leaving open the question whether it is best regarded as singular or plural" (Bowlby, 1980, p. 52).

Though this passage is not a detailed description it does direct research on attachment modeling to the exploration of how different constituent parts for an attachment control system might be organized and integrated. Updating Bowlby's conceptualization can therefore involve posing questions such as: What kinds of subsystems influence attachment behaviour at different ages? How will different subsystems interact? How will the empirical finding of continuity in attachment patterns be supported as higher-level subsystems come 'on-line'? Lastly, how might new subsystems be constructed?

One way to attempt to answer these questions is to design and implement running simulations of possible attachment models (Petters 2004, 2006, Petters and Waters 2015). The field of attachment modeling is in its infancy but there are several examples of

computational attachment models which simulate the action of the attachment control systems. Bischof (1975) presented a highly abstract simulation where the simulated infant agent and carer agent do not have complex internal states nor possess complex perceptual apparatus. They interact in a very simple 2D virtual environment which possesses no other complex objects. The infant agent's social motivation is formalized as a cybernetic control system that is based upon the operation of attachment and exploration behavior systems. The infant agent's behavior switches between security seeking (moving towards the carer agent) or exploration (moving randomly), and the threshold for its behavior switching is moderated by a simple measure of familiarity which does not emerge from interaction with the environment but is manually switched between settings.

Petters (2006) presents agent based models (ABMs) which are more complex than Bischof's in several ways. The infant and carer agents exist in 2D dimensional virtual environments in which they can move about and signal to each other with varying emotional valence. So infant agents model communications which vary from simulated smiling to crying. Infant and career agents also sense the valence of the signals of other agents. A range of other objects that vary in size and functionality are included in the simulated environment. Toy objects have distinct perceivable attributes about which infant agents can learn. These attributes therefore allow infants agents to discriminate between toy objects they have already interacted ('played') with and those that remain unfamiliar and so novel. Infant agent can also become familiar with patterns of large objects, allowing them to distinguish 'home' environments where they have interacted over many simulation cycles and a 'strange situation' environment with unfamiliar objects and arrangement of objects (Ainsworth *et al* 1978). Simulations can also include agents which represent unfamiliar

adults, such as the friendly ‘stranger’ who interacts with infants during the Strange Situation Procedure (Ainsworth *et al* 1978).

In the ABMs described in Petters (2006), the internal state of agents is also much more complex than the agents in Bischof’s simulation. Infant agents possess behavior systems for exploration (of objects); social interaction (with other agents including their carer and a ‘stranger’); fear (of unfamiliarity over a wariness threshold); physical need (for close interaction and food), and attachment anxiety (being far from or out of sight of their carer agent) each of which generate and activate respective goals. Each of these goal generator-activators includes an opportunistic component, in that if an unfamiliar (therefore attractive) toy or social opportunity appears close to the infant this proximity gives rise to high temporary activation in the appropriate behaviour system. In addition, goal activation for some goals has a drive like component (so any time period without exploration, social interaction, or contact with an attachment figure gives rise to a slight incremental increase in activation of the goal activation level for these goals).

As a simulation proceeds goal activation levels for exploration, social interaction, reducing wariness and reducing attachment anxiety goals rise and fall. Several action selection mechanisms have been implemented to decide which goal is the one which sets external behaviour of movement and signaling. A relatively simple mechanism implemented for action selection is a ‘winner take all’ mechanism where the highest current activation controls what goals become active and directs motor and signaling behaviour. Computational experiments have been conducted with infant agents with this kind of control mechanism, where infant agents can learn how effective and sensitive their carer agent is in responding to their signals arising from social interaction and attachment anxiety

goals. These experiments show different patterns of results are produced depending on whether infant agents learn about carer agent effectiveness from bids for all kinds of social interaction, or from just episodes of response to significant attachment anxiety – a theoretically important distinction (McDonald 1992). A more complex control system has also been implemented with dual loci of action selection control – with a winner take all action selection mechanism operating alongside a deliberative subsystem which can inhibit and redirect behaviour after attempting to reason about what the likely outcome of actions directed by the winner-take-all mechanism will be. The deliberative subsystem of this more complex control system includes operators for strategy creation, strategy evaluation and selection, and strategy execution.

Bischof's, and to a greater extent, Petters' simulations should be viewed as early attempts to explore a large design space of potential attachment control system designs. A key insight presented in this paper is the similarity of these kind of attachment control system with contemporary cognitive architectures (Petters 2006, p. 143) suggesting a useful reconceptualization of the attachment control system as a 'cognitive architecture for attachment'. Since Bell and Newell (1971) didn't coin the term 'cognitive architecture' until 1971 it is understandable that this term was not adopted by Bowlby when he originally formulated the attachment control system concept in 1969. However, from a contemporary vantage point, the match is clear. Newell (1990) defined a cognitive architecture as: *"the fixed (or slowly varying) structure that forms the framework for the immediate processes of cognitive performance and learning."* (Newell, 1990, p. 12). So, Bowlby's conception of the attachment control system is certainly a cognitive architecture in this sense. The temporal structure of attachment patterns as long-term control states, and related shorter-

term states in particular attachment related episodes, can be compared with other kinds of affective control states within cognitive architectures (Petters, 2006a; Sloman, 1993). In addition, Anderson (2009) conceptualizes cognitive architectures from three perspectives: their function; the structures and mechanisms they possess; and the computations they perform. Bowlby set out all the three required ingredients for an architectural analysis of attachment phenomena, namely: empirically observed attachment behaviors; information processing structures and mechanisms; and an evolutionary functional analysis that matches these structures and mechanisms with observed behavioral patterns. For example, in the first volume of his Attachment Trilogy, Bowlby did not just match behaviors of interest with potential information processing structures and mechanisms. He also spent a great deal of this first volume concerned with explicating the evolutionary function of attachment behaviors like proximity seeking.

A cognitive architecture does not need to be organized around just four kinds of motivation or goal (attachment, fear, sociability and exploration). Instead, the architecture should be able to support the pursuit of a very large range of goals, and for a complex architecture, pursuing these goals through using various kinds of mechanism or structure. Bowlby presented reflexes, fixed action patterns, planning and Internal Working Models (IWMs) as mechanisms and structures involved in pursuit of all attachment related goals, from infants to adults. Future attachment modeling should update this collection of candidate structures and mechanism and systematically explore the architectural design space of possible cognitive architectures for attachment.

Considering how a modern computational account can take attachment theory forward we can ask: what are the options a modeler might choose from in designing cognitive

architectures? One way to classify different architectures is by the nature of the subsystems they possess. The next section presents the planning systems that Bowlby described as part of the attachment control system, with some consideration of what shape contemporary up-dates might take. The following section presents IWMs as a separate kind of component within the attachment control system, which would work alongside the planning system, and considers how this construct might be updated. However, how the components are arranged in a cognitive architecture is highly influential in determining the architecture's capabilities (Sloman, 1993; Wright, Sloman, & Beaudoin, 1996). For example, Dayan and co-workers have formulated a Bayesian cognition framework, inspired by neuroscience and information theory, that explains how multiple subsystems (they term 'controllers') carry out action selection (Daw, Niv, & Dayan, 2005; Dayan, Niv, Seymour, & Daw, 2006; Dayan & Seymour, 2008; Lengyel & Dayan, 2007). Petters and Waters (2010) show how the different controllers of this framework are putative explanations for a range of attachment phenomena. What is critical about this framework when considering the organization of a cognitive architecture for attachment is that a coherent behaviour pattern might be driven by more than one controller and the responses of the controllers are considered to be mediated by a decision principle based upon uncertainty (Daw et al., 2005; Dayan, 2008; Lengyel & Dayan, 2007). Each of the controllers regulated by this decision procedure incorporates a measure of its own uncertainty in its likely performance. The least uncertain controller is then chosen to direct the next action. This Bayesian decision mechanism may account for the distinction between organized and disorganized behavior because when experiences are unpredictable no controller has enough certainty in its own performance to seize control and direct actions with coherence and consistency.

Hierarchical Planning

Planning involves considering 'in the mind' the outcomes that are possible from taking sequences of available actions. Whilst ethology is mainly concerned with how discrete actions are taken in response to actual immediate stimuli without much 'look-ahead', psychoanalysis has a lot to say about how people think reflect in detail on future outcomes. However, Bowlby's treatment of planning within the attachment control system framework is based far more strongly on the state search approach of Artificial Intelligence planning than psychoanalytic phantasy. In the first volume of the *Attachment and Loss* trilogy, Bowlby also set out differences between control systems in terms of how the behaviors within them are organized. He presented behavioral chaining as an example of a simple organizing principle for control systems, and hierarchical planning as much more complex and flexible (Bowlby, 1969, p. 76). In this approach, plans are composed of sub-plans, and each plan and sub-plan is a set of instructions for action. So, a high-level plan can give a main objective and general strategy, where subplans deal more with the details of how to implement actions.

Bowlby presents a very broad range of planning types. He includes planning examples of both explicit human plans and planning carried out by rats and other animals (Bowlby, 1969, p. 79-80). Although Bowlby presented contexts in which different kinds of plans would be formed, he did not distinguish these planning examples in terms of the sorts of representational or computational details needed to implement running simulations. We should expect modern computational accounts of planning in the attachment domain to situate planning subsystems within an overall computational architecture. Planning systems have been a mainstay of artificial intelligence research from the 1960s to contemporary

approaches that range from planning in robot navigation to planning systems which carry out complex industrial optimization tasks like scheduling the loading of container-ships (Russell & Norvig, 2003). A limitation of traditional AI planning approaches for the attachment domain is they involve considering plans as chains of discrete actions from a fixed repertoire bringing about transitions between discrete states of the world. So not taking into account the continuous nature of relevant variables in close social interactions. Recently, researchers in cognitive science have also implemented systems which are more plausible models of planning in natural systems. For example, Cruse (2003) discusses a neural network model which can be used to control action, but also when minor changes are made the same network can be used to represent the self and environment and carry out planning of actions and outcomes of actions before actions are actually taken. This is an example of a motor system being re-tasked to become a forward model that might be used in planning. Bowlby (1969) certainly considered forward models of this kind for inclusion in the attachment control system:

"the settings to which the servo-units of an anti-aircraft gun are working can be derived from a radar instrument which is so designed that it tracks the aircraft, and not only tracks it but extrapolates from present knowledge of how the aircraft is moving to predict the aircraft's future position. In this way the gun is kept pointing constantly in such a way that a shot fired is likely to hit the aircraft. This type of system also is replicated in living organisms. There is reason to think that our possession of systems of this kind, appropriately linked and integrated, enables us to hit a moving tennis ball, and that similarly linked systems enable a falcon to seize a flying bird. Henceforward the objective of hitting the ball (or the aircraft) or seizing the bird is termed the set-goal of

the system" (Bowlby,1969, p .43)

Grush (2004) provides a broader review of forward models, describing their wide use in diverse theories of human cognition. What all these models have in common is they all emulate the outcomes of actions using forward models of those actions without having to actually take those actions. Grush provides most illustrations for his emulation account from motor control but also outlines other cognitive functions that might be supported by the same forward model emulator mechanisms, including reasoning and theory of mind phenomena. What Cruse's network, Grush's emulators and other contemporary approaches, such as predictive processing accounts (Clark 2013), have in common is an intimate connection between a planning mechanism (that allows an agent to 'try out' actions before actually taking them), and representations of the self and environment that act as internal working models which support the deliberative planning mechanisms - which is precisely what Bowlby set-out in the 1969 first volume of the *'Attachment and Loss'* trilogy (Bowlby, 1969), albeit in less detail and without implemented simulations as examples.

Bowlby presents planning as a bounded cognitive process which is activated within a broader control system architecture. However, hierarchical plans can involve more abstract descriptors at higher levels. Such as descriptors related to the self. Therefore, following Neubauer's (2012) analysis, the products and processes involved in hierarchical planning can be invoked in self-emergence in a less bounded manner. So from emulation of actions to higher level emergence of self, hierarchical attachment planning provides a potentially rich future research direction.

Internal Working Models

The concept of 'Internal Working Models' (IWMs) as important structures within the attachment control system was introduced in the first volume of the *'Attachment and Loss'* trilogy (Bowlby, 1969). Bowlby was very clear about the conceptual heritage of this idea. Unlike his ideas on planning which were introduced directly from contemporary sources, the IWM concept was an attempt to make scientifically respectable previous ideas from psychoanalysis:

"The environmental and organismic models described here as necessary parts of a sophisticated biological control system are, of course, none other than the 'internal worlds' of traditional psychoanalytic theory seen in a new perspective. As in the traditional theory so in the theory advanced, much psychopathology is regarded as being due to models that are in greater or less degree inadequate or inaccurate" (Bowlby, 1969, p. 81)

In his attempt to draw this concept into scientifically oriented (and therefore evaluable) cognitive science, IWMs are described by Bowlby as higher level representational forms which integrate and exert control over lower level control systems. Like psychoanalytic internal worlds, their principal information processing function in the mature theory is to allow predictions to be made about the likely outcomes of taking actions within a given environment.

In the first volume of the trilogy Bowlby emphasizes the requirements for IWMs to be updated. He invokes IWMs at early stages in development and later on, when linguistic skills and conscious reflection can enable models to become more adequate (Bowlby, 1969, p. 84). Bowlby also observes that pathological sequelae of separation and bereavement can be understood in terms of out of date models or half revised models which may contain

inconsistencies and confusions ((Bowlby, 1969), p. 82). Thus, linking to ideas from his three ‘mourning papers’ in the early 1960s where he sets out a three-part process of separation anxiety, protest and repression. In the mature version of attachment theory this idea is no longer described in the terms of psychoanalysis, and repression is instead reframed as ‘defensive exclusion’ (Bowlby, 1980).

Bowlby linked the construction of plans with the operation of Internal Working Models:

“Not infrequently, many alternative plans are concocted, their potential consequences imagined (on the basis of models of environment and organism) and the consequences of each plan appraised. Only after that is any particular plan put into operation” (Bowlby, 1969, p. 114)

As with his presentation of planning, he does not emphasize the representational details for IWMs. This is because his audience at that time was researchers and clinicians in psychology and psychiatry - not researchers in artificial intelligence, robotics or systems engineering. Petters (2016) presents a historical review of the development of the IWM concept in Bowlby’s writing and shows the change over time from an analog to symbolic conception of IWMs between 1969 and 1982.

IWMs have been studied in a variety of contemporary simulations. In the agent based models of Petters (2006) the total set of plans and evaluations of expected outcomes can be seen as a form of IWM. A contrasting way to model IWMs is shown in a number of recent studies using artificial neural networks (ANNs). These ANNs have been used to simulate, in a highly abstract fashion, how attachment representations may change or remain stable over development. Put simply, what the ANNs do is recognize patterns in the input information, and at later point those patterns can influence future ‘behavior’ of the networks, in the sense

of the outputs produced in particular contexts. A number of researchers have used Hopfield ANNs to simulate prototype formation (Edalat & Mancinelli, 2013; Smith, Stevens, & Caldwell, 1999)). Edalat and Mancinelli (2013) present a model that explains early attachment stability as arising from strong dynamic systems patterns. They demonstrate in Hopfield neural networks that simulations starting in a large number of initial conditions end up through multiple iterations in the same final attractor state (so possessing large basins of attraction). Demonstrating the interaction of empirical research and modeling, Fraley (2007) used an ANN to investigate earlier (Fraley 2002) results from a meta-analysis on attachment stability. The ANN simulation demonstrated how early relationship prototypes can be formed and will respond to differing schedules of subsequent experience. A key conclusion from these artificial neural network simulations is that simulations can demonstrate that early prototypes are not over-written, and so show greater continuity, when new relationship experiences are inconsistent. But consistent presentation of new prototypes does result in gradual change.

How might contemporary approaches in cognitive science move this work forward? While ANN models provide one possible way of updating Bowlby's IWM concept, this is just one possible direction from diverse options for future research. The work of Gibson (1986) and more recent sensorimotor approaches to action selection provides a contrasting way to conceive of IWMs for attachment (reviewed in Petters (2016)). Gibson was primarily interested in explaining perception and argued that rather than the function of vision being concerned with gaining information about the properties of entities in the environment, it was better conceived as providing information about what actions are available to individuals - thus giving information about what positive and negative action

affordances are available for the perceiver (Sloman, 2011). We can adapt Gibson's notion of action affordances to the attachment domain - what positive and negative attachment affordances are available for an infant in a particular context? This work is related to existing research looking at how to learn social affordances in human robot interaction (Shu, Ryoo and Zhu 2016). Attachment affordances are more abstract than the social affordances in this study but share with them being non-representational or 'representation-lite' action-focused conceptualization of IWMs. Viewing IWMs in this manner emphasizes that IWMs are not just memory stores but include links from what is perceived to options for actions.

The analysis of Lake, Tomer, Tenebaum, and Gershman (2006) provides an alternative view to both the pattern recognition approach using ANNs to stand in for IWMs, and the attachment affordances approach. Lake et al. (2006) emphasizes the richness and flexibility of human learning and suggest that for human cognition, *"model building is a better metaphor than learning as pattern recognition"* (Lake et al. (2006), p. 38). The basis of this view is that human learners operate in a fundamentally different way to today's ANNs, by using rich prior knowledge when engaging in new tasks. Humans are able to add knowledge to existing models because the models operate as general schemes and intuitive theories. Models can then be constituted as more than 'passive' patterns in networks that only produce output as a direct result of new inputs, but structured procedures that generate new examples of concepts, which can even possess internal compositional structure. Lake et al's(2006) approach gets us closer to Bowlby's suggestion that IWMs transmit, store and manipulate information and allow the individual to *"conduct small scale experiments within the head"* (Bowlby, 1969, p. 81). In this contemporary view, an IWM is a world view, with

internal ‘moving parts’ of the kind that we can acquire implicitly over time or even explicitly wholesale in therapy.

STABILITY AND LABILITY OVER ONTOGENETIC DEVELOPMENT

Bowlby didn’t use the terms ‘precocial’ and ‘altricial’ to describe stable or labile control systems, but the developmental trade-offs he sketched for these contrasting systems match current conceptions in research on altricial and precocial patterns of development in contemporary computational cognitive modeling (Chappell & Sloman, 2007). For example, Bowlby (1969, p. 46) presents a developmental trade-off whereby ontogenetic lability in a control system might result in a longer developmental duration but may also result in this control system becoming better adapted and more flexible than a stable fixed alternative. Recent work has framed similar ideas in terms of altricial and precocial forms of development, within artificial systems which might be used in computational modeling (Chappell & Sloman, 2007). It is notable that this recent work has additionally linked altricial development (which corresponds to Bowlby’s labile developmental pattern for control systems) with the development of higher level more explicit representational forms.

Representational Change Through Ontogenetic Development

According to Bowlby, each behavior system in the attachment control system has its own ontogenetic development, initially producing reflex actions and later in infancy producing fixed action patterns which increase in the complexity of their organization in sequences and chains. Behavior systems related to the attachment system continue to develop through childhood and in adulthood we can use explicit models and natural language to deliberate on the attachment related consequences of our actions. However, according to Bowlby earlier developing ‘behavior system’ processes still exist and operate

alongside more mature systems. Bowlby compared changes in the kinds of representations which infants and children would possess in their attachment control systems with representational forms such as sensorimotor, pre-conceptual and symbolic representations which were theorized to occur in different Piagetian stages (Bowlby, 1969, p. 153). He also speculated that actions which are automatically triggered early in development, such as feeding reflexes, might then be organized into causal hierarchies which can then form part of plans with set-goals. (Bowlby, 1969, p.160; Petters & Waters, 2010, p. 54-55). This appreciation of how representational forms can change early in infancy was matched by an acknowledgement of the important role played by language and symbolic forms of representation in later stages of development:

"Thus, whereas during infancy and childhood humans are incapable of structuring their behaviour in any way more complex than the simplest of plans, in adolescence and adulthood behaviour is habitually structured on the basis of elaborate plan hierarchies. This tremendous development on the sophistication of the behavioural organizations employed is made possible, of course, by the increasing capacity of the growing human child to use symbols, especially language." (Bowlby, 1969, p. 155)

Bowlby recognized that natural language is the ultimate and most sophisticated way in which an individual can represent themselves within their social environment. This form of representation has the benefit that *"instead of each one of us having to build his environmental and organismic models entirely for himself, he can draw on models built by others"* (Bowlby, 1969), p. 82)1. A benefit of non-communicative aspect of language is the possession of language allows more flexible and imaginative plans to be created, and shared with others, with sharing with others constituting a possible form of therapy.

Significantly, Bowlby also recognized that the nature of early representations is intimately linked with later patterns of behavior, for example:

"Man's capacity to use language and other symbols, his capacities to plan and build models, his capacities for long-lasting collaboration with others and for interminable strife, these make man what he is. All these processes have their origin during the first three years of life, and all, moreover, are from their earliest days enlisted in the organization of the attachment behaviour." (Bowlby, 1969, p. 358)

SYNTHESIS: DEFENSIVE PROCESSES, CONSCIOUSNESS AND THERAPY AS META-MANAGEMENT

As with the conceptual development of IWMs, Bowby's treatment of defensive processes shows a strong inheritance from his prior psychoanalytic views. In particular, he transformed ideas on repression into a concept of 'defensive exclusion' which conserved aspects of the older idea but in a new scientifically respectable format that could be integrated with other contemporary cognitive science constructs.

In the final control systems version of this theory of defensive exclusion, he drew for inspiration upon material from the emerging field of cognitive psychology. He explained defensive processes in terms of multiple working models and differences in conscious access (Bowlby, 1973, p. 238); selective attention (Bowlby, 1980, chapter 4); and explained recall, reflection and potential internal conflict in self-image in terms of the distinction between episodic and semantic memory (Bowlby, 1980, pp 61-64).

In his 1973 account, some working models are open to conscious access. Some are unconscious but still highly influential, and so act as *"a version in different terms, of*

Freud's hypothesis of a dynamic unconscious" (Bowlby, 1973, p. 238). Bowlby also suggests IWMs can be formed with incompatible information, some of which becomes dominant and exerts influence either consciously or unconsciously:

"In a person suffering from emotional disturbance it is common to find that the model that has the greatest influence is one that developed during his early years and is constructed on fairly primitive lines, but that the person himself may be relatively, or completely, unaware of; while simultaneously there is operating in him a second, and perhaps radically incompatible, model, that developed later, that is much more sophisticated, that the person is more nearly aware of and that he may mistakenly supposed to be dominant." (Bowlby, 1973, p. 238).

Bowlby also linked the operation of information processing with conscious awareness:

"Reflection suggests that many of the mental processes of which we are most keenly conscious are processes concerned with the building of models, with revising or extending them, checking them for internal consistency, or drawing on them for making a novel plan to reach a set-goal. Although it is certainly not necessary for all such processes always to be conscious, it is probably necessary that some should be so sometimes. In particular, it seems likely that revising, extending and checking of models are ill done or done not at all unless a model is subjected from time to time to whatever special benefits accrue from becoming conscious" (Bowlby, 1969, p. 82).

These quotes show that Bowlby's high-level framework can provide inspiration for contemporary computational models of defensive exclusion. However, although the problem has been analysed no fully implemented simulations of defensive processes have yet been produced. Work towards simulations of defensive exclusion includes Petters and

Coyne-Umfreville (2017) and Petters and Beaudoin (2017) discussing how the differences in self-knowledge demonstrated in dismissing and preoccupied behavior patterns in the Adult Attachment Interview (AAI) might be modelled using contrasting frequency and decay memory recall functions that control what memories become accessible during the AAI.

Part of the challenge in simulating defensive processes is the link to consciousness. When considering how a modern computational approach can take these ideas forward it is worth noting Sloman's observation that current research on consciousness within cognitive science is *"riddled with confusion and muddle"* (Sloman 2011, p. 2). He suggests that this term is better replaced with precisely defined labels for special cases. For example, notions of accessibility or inaccessibility of self-knowledge might be captured in theories of consciousness which invoke something like a reflective or second-order or meta-management processing layer in a cognitive architecture (Lau & Rosenthal, 2011; Minsky, 2006; Sloman & Chrisley, 2003). In this view, considering how an architecture might simulate having thoughts about thoughts is a way to investigate some kinds of self-reflection or meta-management which are of interest to attachment researchers. Of particular pertinence to computational modeling, Bowlby describes how much of our habitual processing is automatic and portrays the kinds of reflective meta-processing on mental life which occurs in therapy (at least when therapy is successful) in explicitly computational terms:

"The psychological state may then be likened to that of a computer that, once programmed, produces its results automatically when activated. Provided the programme is the one required, all is well. [When] *representational models and programmes are well*

adapted, the fact that they are drawn on automatically and without awareness is a great advantage. When however, they are not well adapted, for whatever reason, the disadvantages of the arrangement become serious.

For the task is of changing an over-learned programme of action and/or of appraisal is enormously exacerbated when rules long implemented by the evaluative system forbid its being reviewed. [...] *A psychological state of this kind in which a ban on reviewing models and action systems is effected outside awareness is one encountered frequently during psychotherapy. It indicates the existence of another stage of processing at which defensive exclusion can also take place, different to the stage at which perceptual defence takes place.* (Bowlby, 1980, p. 55-56)

In this passage, what Bowlby is referring to when he discusses therapy as reviewing models are meta-processes such as self-reflection and meta-management (Sloman, 2011). Research focusing on meta-management and self-reflective capabilities operating alongside other action selection mechanisms in complex scenarios such as those found in the attachment domain remains an under-explored area of cognitive systems research (Vernon, 2014, p. 12). Therefore, there is scope for progress in this area, and research can take inspiration from both the requirements set out by Bowlby's account and contemporary modeling approaches.

Cognitive modeling research on memory processes, such as memory accessibility in language use, is much more developed than research on self-reflection and meta-management (Lewis & Vasishth, 2005). However, Donald (2001) provides a analysis that shows the two approaches are related. This is because the differential accessibility of certain kinds of thoughts and memories, the expression of which is observed in attachment

discourse like the AAI, may be investigated by looking at biases and selectivity in memory retrievals. As Hesse (2008) notes, the relatively fast paced AAI protocol has the potential to "*surprise the unconscious*". Particular discourse patterns which reveal a particular kind of internal 'state of mind' occur in part because of what is brought to mind from memory and what is not. Both Petters and Coyne-Umfreville (2017) and Petters and Beaudoin (2017) follow this line of exploration in their analysis of how modeling differential memory retrieval is a way to model defensive processes. Therefore, the current state-of-the-art in attachment modeling does not attempt to explain the first-person phenomenology of conscious self-awareness in the attachment domain. But existing approaches can attempt to explain the routes by which, for some, memories can form part of integrated and coherent narratives, whilst for others, certain memories do not reach self-awareness. So explaining defensive processes in terms of ineffective access consciousness rather than in terms of phenomenal consciousness (Donald 2001).

MODELS THAT BRIDGE SOCIAL AND NEUROSCIENCE LEVELS

Attachment theory investigates a huge range of phenomena - from the secure base behavior of one-year old infants in a public park (Anderson, 1972) to the linguistic narratives adults present about their own childhood experiences (Hesse, 2008). So, attachment modeling also needs to capture meaning-making and attachment discourse, with models which support language and declarative memory. Whilst some perceptual, motor and affective phenomena of interest might be simulated at a low neuroscience abstraction level, this paper argues taking a turn to neuroscience by modeling just at this level is not the way forward on its own. Ballard (2015) sets out a solution to the requirement for attachment models to bridge different levels of analysis. As Ballard notes, understanding the interaction

of high level functions like goal based processing and intentional behavior, and low-level functions such as neural action will likely be more demanding and may even be intractable if the system is only analyzed at lower levels of abstraction. Such as looking at the behavior of individual neurons or low level neural networks. He argues that since the brain has a hierarchical structure comparable to the hierarchical organizing principle of computation, layered structures in the brain can be compared with computational abstraction hierarchies. When models are posited at high abstraction levels, lower level details are initially suppressed but not ultimately ignored. On the contrary, as Ballard notes: *"by telescoping through different levels, we can parcellate the brain's enormous complexity into manageable levels"*. ((Ballard, 2015), p. 18). We can therefore integrate over abstraction levels for domains that require this, like attachment modeling. Having multi-level models can then aid evaluation by bringing to bear stronger empirical constraints because different abstraction levels will all constrain the model.

A benefit of this approach is models with intermediate psychological level constructs can link results from social psychology research on attachment with neuroscience results. This has not tended to occur. For example, whilst an early paper by Hazan and Shaver (1994) launched a social psychology approach in attachment research and this paper did invoke simple control systems diagrams, Crowell, Fraley and Roisman (2016) note that on the whole, the study of adult attachment processes has not tended to emphasize the normative development aspects of the attachment system. Hence the elucidation of the attachment control system in this work has not been central. Rather, according to Crowell, Fraley and Roisman, a social psychology approach in attachment research has mainly focused on individual differences in behavior and cognitions such as expectations. In

addition, from the neuroscience perspective, Coan (2016, p 243) argues that it is important to attempt to bridge currently disparate sub-fields - bringing together research on attachment as a neural construct and individual difference research on attachment. This paper therefore makes a contribution by showing how to integrate lower and higher level approaches through combining both within broad architectural architectural 'control system' models.

CONCLUSION

Attachment theory originated from theories set out by Freud, and other psychoanalysts. However, Bowlby ultimately formed a clear distinction between attachment theory and psychoanalytic theory. He did this by explaining the richness, broad scope and complexity of behavioral phenomena which interest psychoanalysts in information processing terms which are congruent with contemporary cognitive science constructs.

Updating attachment theory by engaging in attachment modeling - that is, using computers to simulate attachment phenomena - has all the benefits that computational modeling is generally found to have in the mostly cognitive psychology research fields in which it is predominantly undertaken (Dawson 2004). A lot of clear thinking has to go into creating running simulations. For theories of psychological phenomena to be run on a computer requires in those theories precision and explicitness in description and a process of formalization which highlights possible logical flaws, inconsistencies, lacunae, hidden assumptions or unexpected complexities about the processes being modelled. In addition to this welcome rigor, the overall process promotes serendipitous discoveries because running simulation can produce unforeseen behaviour (Petters, 2004, 2006a; Petters & Waters, 2015).

A number of specific benefits accrue from reviewing and updating the information

processing foundations of attachment theory. The first is that attachment theory, with its contemporary multifaceted and diverse nature, can be better integrated as a research field. The potential integrative benefits of attachment modeling are underscored by looking at how models can bring together elements from the great diversity of approaches in contemporary attachment theory. A single model can include interacting agents that represent infants, caregivers and strangers, so enabling the Strange Situation Procedure to be simulated (Petters & Waters, 2015). The care-giver software agents that are used to model Strange Situation studies can also be used in simulations of the Adult Attachment Interview (Petters & Beaudoin, 2017; Petters & Coyne-Umfreville, 2017). Having such diverse types of empirical data, from the many empirical attachment measures now used as well as the many perspectives that attachment research now follows, allow attachment model evaluation and validation by constraining models to simulate such a broad specification of requirements (Petters, 2004, 2006a).

Fonagy suggests attachment theory can seem 'method-bound' because of the strong focus on a set of validated measures:

[Attachment theory's] *"scope was determined less by what fell within the domain defined by relationship phenomena involving a caretaking-dependent dyad and more by the range of groups and behaviors to which the preferred mode of observation, the strange situation, the adult attachment interview, and so forth, could be productively applied."* (Fonagy, 1999, pp. 472-473).

While Fonagy's remedy for the method-bound nature of attachment theory is to suggest that attachment theory researchers engage with the clinical discoveries of psychoanalysts, attachment modeling can provide a different route out of what Fonagy perceives is a

method-bound domain. Attachment models can produce new predictions for existing measures and can simulate behavior in contexts not covered by current measures or observations. So while this paper has presented a diverse range of contemporary models as updates for elements of Bowlby's framework, future work can be even broader in scope and ambition.

To sum up, while contemporary attachment theory is diverse and multifaceted with many varied perspectives, most of these perspectives involve collecting new kinds of empirical data not new underlying information processing explanations for attachment theory. Computational modeling that 'telescopes' across abstraction levels (Ballard, 2015) can help update Bowlby's conception of the attachment control system. While modeling can contribute to attachment theory, the benefit is two-way. There is significant value in the rich, deep, broad and complex modeling scenarios that the attachment domain provides for computational modelers.

References

Ainsworth, M., & Bowlby, J. (1991). An ethological approach to personality development. *American Psychologist, 46,* 333-341.

Anderson, J. (1972). Attachment behaviour out of doors. In N. Blurton-Jones (Ed.), *Ethological studies of child behaviour* (pp. 199–215). Cambridge, UK: Cambridge University Press.

Anderson, J. (2009). *How can the human mind occur in the physical universe?* New York: OUP.

Ballard, H. H. (2015). *Brain Computation as Hierarchical Abstraction*. Cambridge: MIT Press.

Bell, C., & Newell, A. (1971). *Computer structures: readings and examples*. New York: McGraw-Hill Book Company.

Bischof, N. (1975). A systems approach toward the functional connections of attachment and fear. *Child Development, 46,* 801-817.

Bowlby, J. (1940a). The influence of early environment in the development of neurosis and neurotic character. *International Journal of Psychoanalysis, 21,* 154–178.

Bowlby, J. (1940b). A parent at hospital. letters to the editor. *The Lancet, June 2,* 704.

Bowlby, J. (1944). Forty-four juvenile thieves: Their character and home life. *International Journal of Psychoanalysis, 25,* 1–57.

Bowlby, J. (1949). The study and reduction of group tensions in the family. *Human Relations, 2,* 123–8.

Bowlby, J. (1951). *Maternal care and mental health*. New York: World Health Organisation.

Bowlby, J. (1956). Psychoanalytic instinct theory. In J.M. Tanner and B. Inhelder (Eds.). *Discussions on Child Development: Volume 1,* (pp. 182–187). London: Tavistock Publications.

Bowlby, J. (1958). The nature of a child's tie to his mother. *International Journal of Psychoanalysis, 39,* 350–373.

Bowlby, J. (1960a). Grief and mourning in infancy and early childhood. *The Psychoanalytic Study of the Child, XV,* 9–52.

Bowlby, J. (1960b). Processes of mourning. *International Journal of Psycho-Analysis, 42,* 317-340.

Bowlby, J. (1960c). Separation anxiety. *International Journal of Psycho-Analysis, 41,*

89–113.

Bowlby, J. (1960d). Symposium on psycho-analysis and ethology ii: Ethology and the development of object relations. *International Journal of Psycho-Analysis, 41,* 313–317.

Bowlby, J. (1961a). Separation anxiety: A critical review of the literature. *Journal of Child Psychology and Psychiatry, 1,* 251-269.

Bowlby, J. (1963). Pathological mourning and childhood mourning. *Journal of the American Psychoanalytic Association, 11*, 500–541.

Bowlby, J. (1969 / 1982). *Attachment and loss, volume 1: Attachment*. New York: Basic Books.

Bowlby, J. (1973). *Attachment and loss, volume 2: Separation, anxiety and anger*. New York: Basic Books.

Bowlby, J. (1980). *Attachment and loss, volume 3: Loss, sadness and depression*. New York: Basic Books.

Bowlby, J. (1981). Contribution to symposium, ‘Emanuel Peterfreund on information and systems theory’’. *Psychoanalytic Review, 68,* 187–90.

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. In I. Bretherton and E. Waters (Eds.), *Growing points of attachment theory and research. Monographs of the Society for Research in Child Development,* Serial No. 209(1-2), 3-35.

Bretherton, I. (1992). The origins of attachment theory: John Bowlby and Mary Ainsworth. *Developmental Psychology, 28* (5), 759–775.

Bretherton, I. (1999). Updating the 'internal working model' construct: some reflections. *Attachment & Human Development* (1) 343-357.

Bretherton, I., & Munholland, K. (2016). The internal working model construct in light

of contemporary neuroimaging research. In *J. Cassidy & P.R. Shaver (Eds.)* H*andbook of Attachment,* (3rd ed.), (pp. 63–91). London: Guilford Press.

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configural altricial information- processing systems. *International Journal of Unconventional Computing, 3(3),* 211–239.

Clark, A. (2013). Whatever next: Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–204.

Coan, J.A. (2016). Toward a Neuroscience of Attachment. In J. Cassidy & P.R. Shaver *(Eds.) Handbook of Attachment,* (3rd ed.) (pp. 242-269). London: Guilford Press.

Cruse, H. (2003). The evolution of cognition - a hypothesis. *Cognitive Science, 27,* 135–155.

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8,* 1704–1711.

Dawson, M. (2004). *Minds and machines: Connectionism and psychological modeling*. Oxford: Blackwell Publishing.

Dayan, P. (2008). The role of value systems in decision making. In ’Better than Conscious? In C. Engel and W. Singer *(Eds.) Decision Making, the Human Mind, and Implications for Institutions.* (pp. 51–70). Frankfurt: MIT Press.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. (2006). The misbehaviour of value and the discipline of the will. *Neural Networks, 19,* 1153–1160.

Dayan, P., & Seymour, B. (2008). Values and actions in aversion. In P. Glimcher, C.

Camerer, R. Poldrack & E. Fehr (Eds.). *Neuroeconomics: Decision making and the brain*'(pp. 175–191). New York: Academic Press.

Dennett, D. (1981). *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA: MIT Press.

Donald, M (2001). *A Mind So Rare: The Evolution of Human Consciousness*. New York: W.W. Norton and Company

Edalat, A. & Mancinelli, F. (2013). Strong attractors of Hopfield neural networks to model attachment types and behavioural patterns. In P. Erdi (Ed.), *Neural Networks (2013 International Joint Conference on Neural Networks)* (pp. 1545–1555). Dallas: IEEE. Retrieved from http://dx.doi.org/10.1109/IJCNN.2013.6706924

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness : A connectionist perspective on development*. London: MIT Press.

Engel, G. (1971). Attachment behaviour, object relations and the dynamic-economic points of view - critical review of Bowlby's attachment and loss. *International Journal of Psycho-analysis*, *52*, 183–196.

Fonagy, P. (1999). Points of contact and divergence between psychoanalytic and attachment theories: Is psychoanalytic theory truly different. *Psychoanalytic Inquiry*, *19:4*, 448-480.

Fraley, R. C. (2002). Attachment Stability From Infancy to Adulthood: Meta-Analysis and Dynamic Modeling of Developmental Mechanisms. *Personality and Social Psychology Review*, *6*, 123–151.

Fraley, R. C. (2007). A connectionist approach to the organization and

continuity of working models of attachment. *Journal of Personality*, *6*, 1157–80.

Gibson, J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates. (Originally published in 1979).

Grey Walter, W. (1956). Presentation: Dr. Grey Walter. In J M. Tanner and B. Inhelder (*Eds.)* Discussions on Child Development: Volume 2, (pp. 21–74). London: Tavistock Publications.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *The Behavioral and Brain Sciences*, *27* (3), 377–396.

Hesse, E. (2008). The adult attachment interview, protocol, method, of analysis, and empirical studies. In J. Cassidy & P.R. Shaver *(Eds.) Handbook of Attachment*, (2nd ed.) (pp. 552–598). London: Guilford Press.

Hinde, R. (1983). Ethology and child development. In P. H. Mussen *(Ed.) Hand book of child psychology* (pp. 27–93). New York: J. Wiley and Sons.

Hinde, R. (1991). Continuities and discontinuities. In M. Rutter (Ed.). *Studies in psychosocial risk: The power of longitudinal data*. (pp.376-384). Cambridge, England: Cambridge Univ. Press. .

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

Lake, B., Tomer, T., Tenebaum, J., & Gershman, S. (2006). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40,* e253.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15,* 365-373.

Lengyel, M., & Dayan, P. (2007). Hippocampal contributions to control: The third way. In *21st annual conference on neural information processing systems* (pp.

1–8).

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing and skilled memory retrieval. *Cognitive Science, 29,* 375–419.

McDonald, M. (1992). Warmth as a developmental construct: an evolutionary analysis. *Child Development, 63,* 753–773.

Minsky, M. L. (2006). *The emotion machine*. New York: Simon and Schuster.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

NIMH. (2016). *National Institute of Mental Health. behavioral assessment - - methods for RDoC constructs.* https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/rdoc_council_workgroup_report_153440.pdf (Accessed:2017-12-14)

Neubauer, R. L. (2012). *Evolution and the emergent self*. Columbia University Press.

Petters, D. (2004). Simulating infant-carer relationship dynamics. In E. Hudlicka and L. Canamero (Eds). *Proc AAAI spring symposium 2004: Architectures for modeling emotion - cross-disciplinary foundations* (pp. 114–122). Menlo Park, CA: AAAI.

Petters, D. (2005). Building agents to understand infant attachment behaviour. In J. Bryson, T. Prescott, & A. Seth (Eds.), *Proceedings of modeling natural action selection* (pp. 158 - 165). School of Science and Technology, University of Sussex, Brighton: AISB Press.

Petters, D. (2006a). *Designing agents to understand infants* (Unpublished doctoral dissertation). School of Computer Science, The University of Birmingham.

(Available online at http://www.cs.bham.ac.uk/research/cogaff/)

Petters, D. (2006b). Implementing a theory of attachment: A simulation of the strange situation with autonomous agents. In D., Fum, F. del Missier, and A. Stocco. *Proceedings of the seventh international conference on cognitive modeling* (pp. 226–231). Trieste: Edizioni Golardiche.

Petters, D. (2014a). Bringing together different pieces to better understand whole minds. In J. Wyatt, D. Petters, & D. Hogg (Eds.), *From robots to humans and back: Reflections on hard problems in the study of cognition.* Cognitive Systems Monographs, vol 22. (pp. 1-7). Cham: Springer

Petters, D. (2014b). Losing Control in the H-CogAff Architecture. In J. Wyatt, D. Petters, & D. Hogg (Eds.), *From robots to humans and back: Reflections on hard problems in the study of cognition.* Cognitive Systems Monographs, vol 22. (pp. 31-50). Cham: Springer.

Petters, D. (2016). An encounter between 4e cognition and attachment theory. *Connection Science*, *28* (4), 387–409. Retrieved from http://doi.org/10.1080/09540091.2016.1214947

Petters, D., & Beaudoin, L. (2017). Attachment Modeling: From Observations to scenarios to designs. In P. Erdi, B. Bhattacharya, & A. Cochran (Eds.), *Computational neurology and psychiatry: Volume 6 of Springer Series in Bio-/neuroinformatics,* (pp. 227–271). Cham: Springer.

Petters, D., & Coyne-Umfreville, E. (2017). Towards modelling adult attachment patterns as control states. In D. Petters, D. Moffat, & J. Parthemore (Eds.), *Proceedings of the international symposium on 'computational modelling of*

emotion: Theory and applications (pp. 177–185). University of Sussex, Brighton: AISB Press.

Petters, D., & Waters, E. (2010). A.I., attachment theory, and simulating secure base behaviour: Dr. Bowlby meet the Reverend Bayes. In A. Sloman, (Ed) *Proceedings of the international symposium on 'AI-Inspired Biology', AISB Convention 2010* (pp. 51–58). University of Sussex. Brighton: AISB Press.

Petters, D., & Waters, E. (2015). Modelling emotional attachment: An integrative framework for architectures and scenarios. In Y. Choe (Ed.), *Neural Networks (2015 International Joint Conference on Neural Networks)* (pp. 1–8). Killarney: IEEE. DOI: 10.1109/IJCNN.2015.7280431

Petters, D., Waters, E., & Schönbrodt, F. (2010). Strange carers: Robots as attachment figures and aids to parenting. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, *11* (2), 246–252.

Richardson, M., Shockley, K., Fajen, B., Riley, M., & Turvey, M. (2010). Ecological Psychology: Six principles for an embodied embedded approach to behavior. In P. Calvo & T. Gomila (Eds.), *Handbook of Cognitive Science: An Embodied Approach* (pp. 161–187). San Diego, CA: Elsevier.

Russell, S., & Norvig, P. (2003). *Artificial intelligence, a modern approach, 2nd edition*. Prentice Hall.

Sloman, A. (1993). The mind as a control system. In C. Hookway & D. Peterson (Eds.), *Philosophy and the cognitive sciences* (pp. 69–110). Cambridge, UK: Cambridge University Press.

Sloman, A. (2011). Varieties of meta-cognition in natural and artificial systems. In M. T.

Cox & A. Raja (Eds.), *Metareasoning: Thinking about thinking* (pp. 307–323). Cambridge, MA: MIT Press.

Sloman, A., & Chrisley, R. L. (2003). Virtual machines and consciousness. *Journal of Consciousness Studies, 10* (4-5), 113–172.

Smith, T., Stevens, G., & Caldwell, S. (1999). The familiar and the strange: Hopfield network models for prototype-entrained attachment-mediated neurophysiology. In D. Franks & T. Smith (Eds.), *Mind, brain, and society: Toward a neurosociology of emotion, vol. 5.* (pp. 213–245). Elsevier Science/JAI Press.

van der Horst, F. (2011). *John Bowlby - from psychoanalysis to ethology: Unravelling the roots of attachment theory*. Chichester: Wiley-Blackwell.

Vernon, D. (2014). *Artificial cognitive systems: A primer*. Cambridge: MA: MIT Press.

Waters, E., Kondo-Ikemura, K., Posada, G., & Richters, J. (1991). Learning to love: Mechanisms and milestones. In M. Gunner & A. Sroufe (Eds.) *Minnesota Symposium on Child Psychology* (Vol. 23: Self Processes and Development), (pp. 217–255). Florence, KY: Psychology Press.

Wright, I., Sloman, A., & Beaudoin, L. (1996). Towards a design-based analysis of emotional episodes. *Philosophy Psychiatry and Psychology*, *3* (2), 101–126.

Footnotes

1 This is a 1969 description which presages Dennett's 1995 description of Gregorian Minds. The similarity may not be coincidental. Dennett termed 'Gregorian Minds' after Richard Gregory. The author of this paper met Richard Gregory and in a subsequent email exchange, in February 2010, Richard Gregory wrote: *" I spent exactly one whole day with John Bowlby when he came to Bristol to visit me and we had a really great*

day. For most of the time we talked about what we were doing in the Brain and Perception laboratory in the Medical School but we did also talk about his work and I have spent quite a lot of time reading his papers and commenting on them. He did indeed have very wide interests and did think in terms of cybernetics and especially interacting machines. I must say I liked him very much indeed, and he certainly had a sharp and imaginative mind."