Towards Modelling Adult Attachment Patterns as Control States

Dean Petters [1] and **Emily Coyne-Umfreville** [2]

Abstract.

Much contemporary research on personality synthesis in computational models and robots is superficial, in the sense that it involves simulating the surface appearance of personality. A deeper conception of personality is that it is a long-term affective control state within a complex control system that is relatively hard to change, slow to learning, and whose causes and effects are more diffuse and indirect than short-term control states like episodes of anger, fear, happiness or relief. This paper will extend this control state conception of personality by broadening it to consider personality as a holistic and multi-level construct. McAdams and Pals' five principles for personality are proposed as an integrating framework for computational modelling of personality as a control state. The paper then argues that extended verbal discourse is a promising modelling scenario to drive the design of integrated multi-level models because it involves multi-level cognitive control and is very demanding in terms of self-reflective meta-cognitive governance. Narrative transcripts from the Adult Attachment Interview (AAI) are then presented as a valuable source of detailed empirical data to be used in creating modelling scenarios across lifespan development. Lastly, the paper presents some initial mini-scenario elements and a sketch of an architecture to be developed to simulate these mini-scenarios.

1 PERSONALITY, ATTACHMENT AND CONTROL SYSTEMS

In 1995, in the paper *"What sort of control system is able to have a personality"*, [28], Sloman suggested that personality should be viewed as a long-term intelligent control state. In this control systems view, high-level and long-term personality states are comprised of a collection of component lower level and shorter term sub-states that act together in a coherent manner. They act as control states by initiating, maintaining, modifying and terminating information processes related to desires, beliefs, deliberations about the consequences of actions, intentions, plans and other substates in mind-like control systems ([27], p. 10). Shorter-term affective control states such as anger have functions such as communication and action readiness. For example, if an organism perceives a threat, a global alarm may be triggered that momentarily 'hijacks' the entire architecture before returning to the initial state. Behaviour characteristic of a particular personality is produced in hierarchical control systems because the system's high level control states are causal control states. Higher-level control states manage lower control states, including emotional states, and through these manage the system's interaction with the world. High level causal states can be functional in the sense of being adaptive, effectively managing overall behaviour in a coherent manner. Alternatively, some high-level personality control states may be dysfunctional or even pathological in terms of how effectively they direct behaviour.

The issue of how to model personality control states is relevant to attachment modelling because both personality and attachment states can both be viewed as long-term control states composed of shorter term control states. Figure 1 presents both personality types and attachment states as long-term control states which are related to other types of semantic control state. However, the fact that both personality and attachment are long-term control states does not mean that personality and attachment states are similar in all respects. Importantly, personality is highly inheritable [7]. In contrast, in attachment research, inherited temperamental traits have been found to be orthogonal to attachment classifications. For example, attachment state is relationship specific - individuals may have different attachment classification with different caregivers [30]. So to summarise the comparison of attachment status with personality - whilst attachment status is learnt, it still comprises a long-term control state within a complex control system.

LONGER TERM	INTERMEDIATE	SHORTER TERM
Personality, Temperament, Attitudes, Skills, Emotions such as love, grief, *Attachment states*	**Moods, Beliefs, Preferences, Emotions such as joy, fear, Intentions, Plans, Desires**	**Neural and physical events,**

Figure 1. Classes of semantic control state, which are compared with respect to the approximate duration that each class of control state may exist as a disposition within an architecture (adapted from [28] and [17]).

1.1 Computational models of Personality

Research papers based on viewing personality as a long-term control state are not readily evident in the contemporary literature on computational modelling of personality. In their 2014 survey of this field, Vinciarelli and Mohammadi organise the review of personality computing research in three categories:

- Automatic Personality Recognition, which involves making inferences of the true personality of an individual from automatically

[1] Birmingham City University, UK, email: dean.petters@bcu.ac.uk
[2] Birmingham City University

processed behavioural evidence. The evidence can include distal cues such as written texts and logged information from electronic devices;
- Automatic Personality Perception, which involves making inferences of personality based on automatically observable behaviours, such as speech, other aspects of paralanguage, nonverbal behaviour, and communication from social media;
- Automatic Personality Synthesis, which involves generation of artificial personalities through software and embodied agents including robots

Whereas automatic personality recognition and perception are directed towards creating systems for classifying human personality, automatic personality synthesis involves creating artificial systems like software agents or robots whose behaviour can itself be recognised and perceived as having personality. However, of the nineteen papers which Vinciarelli and Mohammadi cite as examples of automatic personality synthesis, none involve simulating personality as a long-term control state in the manner described by Sloman [28]. There are some examples of less superficial implementations. Surendran and Long do provide an example implementation of a system from which temperament-like states emerge [29]. However, this highly abstract simulation is not intended to closely match human temperament. So overall, what contemporary work in this field fails to provide is richer theories and models which explain the causal and functional interactions between a variety of long-term and shorter term motives, and related control states.

2 PERSONALITY AND PERSONAL NARRATIVE

McAdams and Pals [14] put forward a comprehensive personality psychology framework for understanding the whole person as an individual. To do this they draw together five principles. for an integrative conception of an individual. We will review these five dimensions from the perspective of the design based approach [31]. That is, considering how the five principles might be simulated together in integrated systems. In McAdams and Pals' view a holistic approach includes:

- **considering human behaviour as more or less constrained by evolutionary design**. Much of the nature humans possess in common in personality terms results from our shared possession of species specific designs adapted to a cognitive niche [26]. Human needs from nutrition and sleep to attachment, autonomy, competence, relatedeness should all be explained, at least in part, by this approach. For the computational modeller, this approach requires architectural designs to be based on some fixed core architecture which is provided by evolution as a starting point for lifespan adaptation and customisation to the environment an individual experiences.
- **the five-factor psychometric model of personality**. This organises personality traits into categories labelled: extraversion/introversion, neuroticism, conscientiousness, agreeableness, and openness [8]. The approach of McAdams and Pals described here subsumes this five factor psychometric model as just one principle of a broader view of personality. Traits are reasonably good predictors for life outcomes like work and relationship success [2]. Researchers have speculated that the kinds of questions that would be expected to arise in social groups in the environment of evolutionary adaptedness are addressed in a particularly apt manner by the five-factor traits [7]. For example, extroversion/introversion can be associated with social dominance; neuroticism with stability; agreeableness and conscientiousness with different aspects of potential for cooperation; and openness with capacity for change and learning ([14], p. 208). However, dispositional trait dimensions are decontextualised and general - speaking to an overall style of engagement with life. We can see from Vinciarelli and Mohammadi's recent survey that this principle forms the framework for most personality computing. A design based approach would simulate the emergence of these behaviour patterns by complex internal functional processes.
- **characteristic adaptations which are more contextualised than traits**. McAdams and Pals suggest that this level of description for personality is of interest to psychotherapists, counsellors, mentors, life-coaches and parents because it focuses on questions like: *"What do people want? What do they value? How do people seek out what they want and avoid what they fear? How do people develop plans, goals, and programs for their lives? How do people think about and cope with the challenges they face? What psychological and social tasks await people at particular stages or times in their lives?"* ([14], p. 208). The characteristic adaptations which McAdams and Pals mention as underlying behaviour related to these kinds of questions is similar to the kinds of intermediate-term control states that Sloman and coworkers [28, 31, 17] present. A combined list from these sources includes: motives, goals, plans, strivings, strategies, standards, values, virtues, attachments, preferences, attitudes, moods, ambitions, obsessions, grief, desires, intentions, schemas, self-images, emotions like fear, anger, happiness and relief, and mental representations of significant others.
- **life narratives that are the integrative life stories that are constructed by individuals in a process of meaning making and personal and social identity formation**. Life narratives can incorporate the reconstructed past, imagined future and allow an individual to *"keep going"*([14], p. 209) through biasing the perception of events so that they become assimilated into a *"more or less coherent whole"*([14], p. 209). Self narrative can augment the predictive power of personality measures beyond those of parametric trait measures. For example, hopeful endings in self-narrative predict future well-being. Whilst each life story is unique, emotional tones in narratives cluster in life narrative themes. For example, a life narrative labelled 'the redemptive self' is a kind of life story about being delivered from suffering which is linked to individuals who become productive, caring and prosocial ([14], p. 210). Narrative interpretation of experiences can prime future growth. Positive examples include: *"I found out how to make our relationship better"* and *"I hope that never happens again"* ([14], p. 210). Narrative accounts that include exploration and accommodation moderate the dispositional traits of openness. Computational modelling of life-narratives is clearly a major challenge. However, an important contribution of this paper is to argue that behavioural scenarios can facilitate modelling this sort of phenomena when they include particularly rich behaviour patterns that arise in systematically structured contexts (see section 2.2).
- **considering how culture affects different levels of personality in different ways**. It can do this by providing a palette of scripts, plots, role-models, warnings, tokens, and images for an individual when they are constructing their narrative. As with life-narratives, modelling an individual's embedding in a culture is a major challenge. However, the benefit of attempting this sort of modelling would be the ability to computationally model phenomena like so-

cial marginalisation and radicalisation as deep internal processes with latent variables rather than through statistical analysis solely focused on externally observable behavioural variables.

2.1 PERSONALITY AND CONVERSATION

If computational modellers accept the grand challenge of creating simulations which integrate across McAdams and Pals' five principles they need a source of rich and detailed scenarios to drive model design and allow model evaluation. Donald illustrates the requirements for multi-level cognitive control using the example of a group of people deeply involved in a lengthy conversation [9]. The challenges of 'keeping up one's end of conversation' illustrates the demands of processing the interchange of ideas and opinions and making appropriate contributions in an ongoing conversation. Social events can be demanding to 'hold in mind' as they can take time to unfold. People in conversation generate novel, rich, and meaningful material, highly changeable as the conversation shifts - and this ability to respond appropriately to novel material, and produce novel but relevant material in return, can extend over an extended period of time. Donald argues that this requires continuous self- and other-monitoring over multiple time scales which is very demanding in terms of self-reflective metacognitive governance and the requirement to store large amounts of knowledge [9].

Conversations can extend from minutes to hours. During these episodes what is said and thought is converted to long-term memories, which continue to causally influence ongoing interaction. What is said or thought at one moment in time can gain momentum and change the overall direction of conversation. Any conversational episode has a broader physical and social context into which it is situated and bracketed. When participants stop taking this into account they are in a sense, removing themselves from the reality of the situation. In this view, effective meta-cognitive processes keep participants tied to reality, and lapses in these processes can leave participants 'in a world of their own' ([9], p. 50). To keep up, each person involved must track what is said and self-monitor thought nearly continuously. Many of the behaviours that enable participants to behave in line with the context will be directed by unconscious scripts or schema. This all means that conversation requires prodigious skills in dynamic memory organisation, in accessing memories appropriately, and in storing new facts for subsequent use in that conversation or later. In this view, personality states are partly constituted of biases and predispositions in cognitive control [9].

To explain this feat of multi-level awareness and governance, Donald proposes a tripartite model, with: (1) momentary binding; (2) short-term control; and (3) intermediate and longer term mechanisms of awareness and governance. Perceptual binding explains the raw feeling awareness. Donald terms this level 1 awareness and governance. Short-term awareness is what is measured in laboratory experiments in consciousness research in experiments that typically only last a number of seconds. Donald terms this level 2 awareness and governance. At this short-term level of awareness and governance, controlled processing arises from the operation of short-term working memory which extends perception to capture simple events which can possess multiple active foci[3]. Working memory affords the capacity to hold an image or memory in awareness and so can allow responses to be delayed. Working memory in its simplest form is a kind of storage that sustains a perceptual trace in the absence of the stimulus that produced it. Attention acts as a 'gate-keeper' to working memory, which is also involved in controlled processing of evaluations, selection, problem solving and response choice ([9], p. 186). Donald's level 3 intermediate and longer-term awareness and governance is extended by the operation of working memory through 'fast switching' of information in processing buffers in and out of long-term memory. Lewis and Vashishth's [12] simulation of sentence comprehension is an example of intermediate control as parsing of garden path sentences is carried out with a small capacity for working memory and mechanisms for 'fast switching' between production buffers and declarative memory. Intermediate awareness and governance extends further into minutes, hours and days through deeper integration with long-term memory which includes the kinds of characteristic adaptations listed in section 2. The kinds of 'events' in level 2 and level 3 governance can be distinguished not just temporally but also by the breadth of level 3 governance to include broader social, cultural and self-concerns [9]. This means at the third level of awareness and governance control states like goals and plans direct ongoing behaviour, and these are influenced by other control states like standards, values, and preferences.

The level 3 system includes voluntary movement and self-initiated actions, as well as supervisory evaluative processes. Particularly human self consciousness emerges from level 3 governance. According to Donald [9], level 3 awareness and governance supports complex states of integration with the social and historical environment which Donald terms 'deep enculturation':

> *"When broken down into their components, the skills we acquire from deep enculturation can be reduced to chains of algorithms that can control attention and emotional valences. Attention determines the sequential flow of memory fixations and perceptual comparisons, and these determine the precise quality and sequencing of subjective experience, producing unique juxtapositions in the mind's eye and influencing what habits we form and interpretations we place on events. The emotional valences attached to various objects, events, and people are an important part of the same process of conditioning the conscious mind. Such algorithms establish the continuity of experience. There is a coherence, an interconnnectedness, about conscious experiences that makes them very different from unconscious ones, where ideas and images can coexist in a pell-mell disorganised manner"* ([9], pp. 212-213)

According to Donald, deep enculturation arises from humans developing from infancy submerged with symbolic cultures in a radically different process from any other species. Taken together, levels, 1, 2 and 3 afford episodic awareness of elaborate event representations. Each episode comprises multiply bound percepts which are chunked into coherent representations. Deep enculturation refers to processes beyond individual episodes to cite and fundamental structuring of the mind. Humans acquire symbolic skills and concepts supported through symbolic skills from the outside in [9]. Some of the these concepts are standards and values which organise behaviour in a general sense. From the perspective of modelling long-term control states like personality we can see it is perceptual, affective and

[3] During these episodes what is said and thought is transferred into long-term memories, which continue to casually influence ongoing interaction. Traditional models of memory propose working memory is simply a transient mode through which the information was encoded into long-term memories [3]. Ericsson and Kintsch [10] propose that in order to carry out skilled activities an addition of a long-term working memory (LTWM) in modelling would explain the ability to leave a task and return to it without impaired performance. This LTWM could explain the ability to have continued fluent conversations with the ability to recall post event.

other processing biases that influence the 'fast switching' of information into and out of working memory that is important. Different personalities will be constituted of different patterns of perceptual and memory retrieval biases and action predispositions. Petters covers similar ground when he discusses how control states 'move' around an information processing architecture:

> *"there is constant relocating and transforming of motivators which is termed circulation.* [...] *useful control states become more influential and 'percolate' up a hierarchy of dispositional control states. Ineffective motivators wither away in influence. One important process is 'diffusion', in which the impact of a major motivator leads it to become gradually distributed in myriad control states which can include new motive generators, plans, preferences, predictive models, reflexes and automatic responses* [...] *Meta-management attempts to influence these numerous processes but some are more controllable than others."* ([20], p, 39)

To model personality-like states we need to implement processes such as these. Information processes that bridge the moment to moment operation of working memory with much longer term relocation and transformation of longer term motivators are what lead to the emergence of personality control states [4]. The next section presents results from attachment research, a more helpfully circumscribed domain than personality. Whilst attachment phenomena range across the lifespan they are focused on issues of responsiveness, sensitivity, predictability and trust in close relationships.

2.2 USING VERBAL BEHAVIOUR TO CLASSIFY ATTACHMENT STATUS

Whilst open ended unconstrained interviews or conversations may tell us a lot of details about an individual, we cannot then readily compare those details with other people in a systematic fashion. If we look at individuals in different contexts, it is difficult to disentangle the influences arising from within the person and the shaping influences of the particular situation they are in. Whereas using standardised psychological procedures designed to assess an individual's state of mind allow easier and more meaningful comparison between individuals.

The Adult Attachment Interview (AAI) [11, 22] is not a kind of free-wheeling conversation that might be taken in any direction whatsoever, such as the kind of conversation which Donald describes as a challenge for multi-level awareness and governance (see section 2.1). It is an interaction between an interviewer and interviewee which has the objective of classifying an adult's emotional and social attachment status through interpretation of their verbal behaviour. The AAI is designed to elucidate structural variations in how life history is presented that allow reliable inferences about the participant's internal state with regard to attachment [11]. The AAI possesses a pre-specified interview format of 20 questions in a fixed order, but with additional specific follow up probes to these main set questions. The questioning and following arrangement must only highlight but not alter the participant's natural response tendencies. It opens with a question asking for a general description of family relationships in the speaker's childhood. Further questions are asked about separations and experiences of rejection; and the effects of these experiences on adult personality. A key section probes experiences of bereavement. Experience of abuse is also asked about. The AAI ends with the speaker being invited to express wishes for his or her real or imagined child in the future [11]. The relatively fast pace of the AAI means it acts somewhat like a mild 'stress test', bringing forward an individuals state of mind with regard to attachment:

> *"The central task the interview presents to participant is that of (1) producing and reflecting on memories related to attachment, while simultaneously (2) maintaining coherent, collaborative discourse with the interviewer (Hesse 1996). This is not a easy as it might appear, and George and colleagues (1985, 1996) have remarked upon the potential of the protocol to "surprise the unconscious". As indicated above, the interview requires the speaker to reflect on and answer a multitude of complex questions regarding his or her life history, the great majority of which the speaker will never have been asked before. In contrast to ordinary conversations, where the interviewee has time for planning, the AAI moves at a relatively rapid pace, and usually all questions and probes have been presented within an hour's time. Ample opportunities are thereby provided for speakers to contradict themselves, to find themselves unable to answer clearly, and/or to be drawn into excessively lengthy or digressive discussions of particular topics"* ([11], p. 555)

The AAI uses adherence to (or violation of) maxims for discourse coherence as a proxy for how an individual thinks and feels about their attachment experience. Transcripts with a lack of overall coherence end up being categorised as such due to major contradictions or inconsistencies, passages that are exceptionally short, long, irrelevant or difficult to understand and follow. The coding for the AAI considers the use of language rather than making retrospective inferences about the person's actual attachment history [11]. It is not what actually happened to an individual in their past that is important for predicting an adult's attachment approach, but the coherence of the attachment narrative that the adult produces in the constrained AAI. So adults of all AAI classifications may publically state the same kinds of values.

In the AAI, interviewees can be classified into four categories. The secure, dismissing and preoccupied enmeshed patterns are the most popular adult classifications and considered 'organised' responses:

- secure autonomous adults express value for attachment relationships and experiences, and give apparently objective responses when asked about any particular relationship experience. When reporting specific experiences they provide confirming detailed memories and demonstrate ability to reflect on those experiences with an understanding of why they, and others, behaved the way they did - and this is the case for happy and troubled experiences. So a secure autonomous adult might describe episodes of rejection but recognise the limitations of attachment figures in a balanced way, as well as include positive aspects of inadequate attachment figures. So compared with other classifications, only secure autonomous individuals are able to access all memories and respond to queries about those memories in a controlled and appropriate manner.
- dismissing adults devalue, or are emotionally cut off from attachment relationships and experiences. These individuals provide very short transcripts, with little to say about specific incidents and attachment experiences from their childhood in general. Responses are not only short but minimise the importance of relationships in general. They may idealize relationships as 'loving' but not provide detailed examples to justify such positive summary statements. Compared with other classifications, dismissing

show attenuated access to memories, bias in the memories that are provided, and an avoidance of the conversational subject of attachment.

- preoccupied/enmeshed adults are preoccupied with (enmeshed by) early attachments or attachment related experiences. When reporting experiences these adults give plenty of detail and long transcripts but fail to provide a good overview because they become so entangled in the details. They may seem to be still engaged in emotional struggles related to attachment relationships Compared with other classifications, preoccupied/enmeshed individuals have access to past memories, but show bias in the memories that are reported, and show a lack of control in how these memories are reported.

The unresolved/disorganised classification is less frequent than the organised categories in non-clinical populations:

- unresolved/disorganised adults speak in unusual ways about loss experiences, and exhibit: interruptions to cognitive processes, particularly in contexts associated with the lost person; disbelief that loss has occurred or is permanent; unfounded fear of death; incomplete mental and behavioural search processes, disorientation in contexts linked to the lost person; and major lapses in metacognitive monitoring of reasoning and discourse processes.

Longitudinal studies have looked at the relationship between current AAI classification for adults and previous categorisation of their behaviour as infants many years previously. Waters and co-workers showed that 72% of participants had the same secure or insecure classification in infancy and adulthood [30]. They found that high levels of life trauma had an impact on the AAI data and when those participants were removed the AAI could predict 78% of attachment security. Main and co-workers have presented similar results [13].

The adult's internal 'state of mind' as indicated by coherence with respect to past attachment relationships is the best predictor of how they will conduct future attachment interactions, not the actual nature of their previous attachment interactions or their explicitly professed values. In addition, there is evidence that the state of mind of an infant's parent has a critical impact on the state of mind for that infant as he or she develops on through childhood, adolescence and into adulthood. As Bretherton and Munholland note:

> *"Overall, AAI findings suggest that parents induct their infants into a way of relating that is consistent with their own secure or conflicted/defensive models of self in relationships. Developmental continuity from nonverbal behavioral and emotional attachment patterns have been established in several longitudinal studies of middle-class families, but is great for security than for specific subtypes of insecurity"*([5], p. 118)

Longitudinal studies also provide valuable detailed evidence of what aspects of early environment bring about later attachment classification. Broussard and Cassidy found that *"adult participants whose mothers had held negative perceptions of them as newborns would be more likely to be classified as insecure on the AAI than participants whose mothers had held positive perceptions of them"* ([6], p. 159). This association was gained for participants whose own AAI was measured between 27-43 years after their parent's perception of them as a newborn infant with a projective measure that involved their mother comparing them with "an average baby". So it demonstrates a significant association with a parent's state of mind and attitude towards their newborn and the AAI classification many years later of the grown-up infant. For the computational modeller, these associations between individual difference categories, early caregiving environment, expectations and values provide valuable constraints to evaluate and validate attachment models. The modelling effort can start by re-producing response patterns from the AAI, and designing architectures to produce these linguistically mediated interview responses. Modelling can then go on to show how the same architectures can produce different caregiving response patterns, and hence demonstrate empirically observed patterns of intergenerational transfer [22].

3 A PRELIMINARY ARCHITECTURAL DESIGN

Simulating patterns of verbal interaction in the AAI will require a complex architecture. This section presents a simple architecture which falls short of this challenge but which can show the direction of travel for the modelling effort. The hybrid architecture illustrated in figure 2 situates reactive subsystems alongside a deliberative planning subsystem (that allows 'look-ahead' reasoning) and a simple meta-management subsystem (where cognitive meta-processes operate on other cognitive processes) [17, p. 103-151]. In this hybrid architecture, the attentive processes that occur are those not stopped by a resource limited variable attention filter. These resource bound serial deliberative processes take input from non-attentive reactive or perceptual processes which operate in parallel. Reactive motive generactivators are triggered and activated by any possibly relevant internal and external events. In the attachment domain there will be possible threats but also possible exploratory and social opportunities. When these conditions are met a motivator is constructed which may 'surface' through the attentional filter and be operated upon by processes in the deliberative or meta-management levels. Amongst the deliberative attachment processes generated by motivators are the creation, selection, and execution of action plans. Deliberative processes that evaluate other processes occur in the meta-management layer. So an agent operating with this architecture can perceive the world, record events, retrieve memories of events and plan future action plans. What the architecture cannot do at all is communicate through language with other agents. So this architecture might be augmented with mechanisms such as those presented in Lewis and Vashishth's [12] simulation of sentence comprehension alongside mechanisms of language production.

4 DEVELOPING SIMULATION SCENARIOS IS AN ITERATIVE PROCESS TO GET THE ABSTRACTION LEVEL RIGHT

Modelling the AAI to simulate attachment behaviours is more challenging than modelling infant attachment behaviours due to the complex involvement of language in these interactions. Infant attachment behaviours can be simulated in terms of proximity to location of carer, orientation and non-linguistic communication [15, 16, 17, 18, 22]. The fact that linguistic interaction is central to the AAI means that cognitive processing such as memory recall, speech planning, considering the likely responses of others, and self-reflection, needs to all be included and integrated. This means that how social standards interact with other motivational or affective states is also going to have a more significant impact in AAI modelling. In terms of training, the agent is going to require a lexicon including attachment related words as well as grammar that can simulate and capture typical AAI responses, but simulating the complexity of full natural lan-

Transcript fragment from real AAI	**Abstract description**	**Description framed in scenario ontology terms**
My mom would stick up for me to the teacher, or to a kid's parent, or anybody, really. I could put it another way, too. I just knew where I stood with her, and that she'd be comforting if I was upset or crying or something. Oh, maybe you wanted a specific example. Um, that time I set fire absolutely positively wasn't supposed to use... came running when the neighbours phoned the fire department about the smoke. I expected to get the life lectured out of me, but she just ran straight for me and picked me up and hugged me real hard. Guess she was so scared and glad to see me, she just forgot the lecture	Specific evidence used to support the statement relying on distinct memories. No violations of relevance, stays on topic. Flowing discourse, particularly when engaging in memory recall [In the process of scenario formation entities such as chemistry sets, smoke and fire department are abstracted away]	**Infant 'training' phase** Carer agent defends infant agent from other agents; comforts and supports during anxiety or distress; responds sensitively when there are cues to danger. Infant agent seeks proximity in balance with exploration. **Adult 'test' phase** Clear distinct memories that are verbalised fluently. Agent is able to give clear examples from infant training phase. Responses have high level of relevance to question.
I don't remember ... (5 second pause). Well, because she was caring and supportive. [interviewer prompt] Well ... (5 second pause), I guess like, well you know, she drove me to school, and I was always really proud of her, I mean, she was really pretty, and she always took care of her appearance	Not convincing support for adjective chosen. Attempt to create a positive picture. Response brief and broken in nature. [In the process of scenario formation entities such as cars are abstracted away]	**Infant 'training' phase** Carer agent provides basic protection but is predictably (reliably) less responsive and less sensitive. Close proximity is a less high quality close coupling and cues unease. **Adult 'test' phase** Memories from training phase are not recalled in detail. Responses are short and conform to externally provided social standards
Uh, yeah, sort of very loving at times, like people were in the old days- uh, my youth, lot of changes since then. I remember home, and home was good and that. And uh, loving, my wife is loving with [child] - taking him out to the movies tonight, special thing he's been wanting to see all week, dadadada. Actually, it's been more like a month, that turtle movie, don't like it too much myself. Too many turtles- where are they from, outer space? Saw it, though, now, when was it, um, maybe 6 months ago. Yeah she's very loving with [child]. really special, really grateful to her for that. my childhood, I remember just sitting on the porch, rocking, rocking back and forth, watching my parents, or maybe having some lemonade- or, you know, this, that, and the other. special sorts of things, just me and her. I wasn't easy, my temperament was hard on her, kind of hard. Me and my cousins from [Town] going down soon - really big birthday, she gonna be 80, gives my age away (continues)	Agent unable to stay with the question. Agent moves to irrelevant topics and memory for events is an issue. Speech is vague and comprehension is poor. [In the process of scenario formation entities such as movies are abstracted away]	**Infant 'training' phase** Carer agent provides basic protection but is unpredictably (unreliable) in how it responds and how sensitive it is. Close proximity does result in a high quality interaction. **Adult 'test' phase** Incomplete marginally relevant responses. Longer responses, both in number of utterances and also in terms of number of not recalled.

Figure 3. Example mini-scenarios using quotes by Hesse [11]. Question 3 in the AAI asks the participant to *"Think of five adjectives, words, or phrases that would best describe your relationship with your mother during childhood- say, between ages of 5 and 12, but even earlier if you can remember. Take a minute to think, and then I am going to ask you why you chose them."* This question involves two levels of processing, the linguistic semantic description as well as memory to recall the events associated with this. Column one of the table shows three example quotes from interviewees who all used the descriptor 'loving' but who were classified differently because of the discourse properties of their full responses. Column two is an abstraction of the response pattern in column one. Column three describes a mini-scenario element which captures the abstraction pattern in column two but presents it in the context of the scenario ontology.

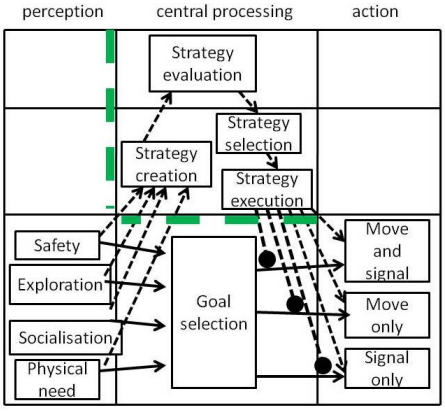

Figure 2. A hybrid attachment architecture with reactive, deliberative and meta-management subsystems. The resource-constrained deliberative subsystem takes input from the reactive-subsystem, carries out 'look ahead' reasoning, and can inhibit the reactive subsystem and execute alternative actions. The green dashed line represents the fact that in the human attachment system deliberative and meta-management processes require attention and so are resource bound, which limits the number that can be concurrently active. Currently, only a simple form of communication is implemented in this architecture, with the agent able to receive and send communications signalling affective tone. However, since the architecture does allow internal processing of plan representations an extension for the architecture will include adding the ability to broadcast and perceive these representations.

guage processing will be avoided. Part of the rationale for modelling attachment is that close inter-personal interactions can be described in a more circumscribed manner than less close social interactions within large groups. So the language level of interaction in the simulation will only be complex enough to capture abstract attachment interactions and not the richness of real AAI transcripts [22].

In the first instance, to model AAI responses the three main classifications must be abstracted; secure autonomous; dismissing; and preoccupied/enmeshed. Scenarios will include the discourse and attentional flexibility that is typical for each of these classifications. However, whilst capturing patterns of discourse, the simulations will abstract away many concrete aspects of the real world that human participants refer to. For example, secure autonomous agents will be able to retrieve memories and reflect on them in an organised way but the actual objects or episodes referred to will be abstractions. Similarly, whilst dismissing agents will express memory biases and an avoidance of attachment related discussions the biases and avoided topics will be in terms of the abstract simulation ontology. The same process of abstraction will occur in scenarios for preoccupied/enmeshed attachment. However, dismissing scenarios will required an opposite kind of memory bias to dismissing scenarios, with a lack of control over the reporting of memories that leads to more emotionally charged memories being recounted [22]. Furthermore, physiological responses to infant cries have been linked to AAI. Ablow and co-workers [1] looked at the response in expectant first time mothers. Those classified as secure on AAI showed physiologically calmer reactions to the crying than those categorised as insecure. When infant sensitivity was tested at 9 months postpartum the secure group had higher levels of infant sensitivity.

Figure 3 shows a scenario fragment contrasting actual AAI transcript fragment, a highly abstract description, and then a more concrete description couched in terms of the scenario ontology (which is significantly more abstract than the actual transcripts but allows for some concrete operational descriptions of phenomena at this abstract level). Producing scenarios that capture these behavioural patterns will be an iterative process that find the right level of abstraction for the scenario ontology [17]. This ontology must possess enough details to allow the AAI behavioural patterns to be simulated but also abstract enough for the simulation to be tractable and not get bogged down in detail which his not relevant to the phenomena of interest. Future work will involve taking fragments of scenario such as those presented in 3 and combining them with other mini-scenarios to gain an overall specification of requirements which is representative of AAI responses generally. Producing scenarios which gain a comprehensive (if highly abstract) coverage of the behavioral domain is important for model evaluation and validation. Whilst cognitive models based on simulating timings or response accuracies can be quantitatively evaluated, this is not possible for models which are not attempting to simulate this kind of data. So evaluation and validation can be carried out by seeing how particular architectures from an architectural design space manage in simulating a broad range of competencies.

A general pattern for scenario elements is to describe events which occur in the 'infancy' stage of the agent simulation, where agents memories are recorded. Then in the 'adult' stage of the simulation that agent is asked questions, accesses memories (in a more or less effective manner) and a response pattern in verbal behaviour is produced. Another way of saying this is that agents that represent caregivers and infants in a early experience 'training' stage for the simulation, the infant agent becomes an adult agent in the AAI component of the simulation, and it is asked questions which require it to draw upon its recorded memories of its infant experiences. In these simulations the contribution to knowledge is not intended to be the sophistication of the memory representations or linguistic utterances but the way that particular agent architectures model processes such as memory recall being effective or ineffective because of the manner in which either memories become inaccessible or the opposite occurs and disturbing memories are recalled even when this is not relevant. So model evaluation and validation will be in terms of how well the simulations reproduce defined qualitative patterns of behaviour.

5 CONCLUSION

This paper presents the case for using rich empirical data from AAI studies to drive model formation of long-term personality-like control states. In this view, attachment styles which are observed in psychological observations such as the AAI arise from the operation of internal control states which are formed from the past experience of low level events and activation of particular short-term control states. In addition, possession of a certain attachment style predisposes an architecture to particular short-term activation patterns for states such as plans or emotions in the future. The function of long-term affective states like attachment style is to organise past and future behaviour at a higher level. In Donald's terms, longer term awareness and governance by an attachment control states oversees processes of moment to moment behaviour from a 'larger landscape' well beyond the immediate perceptual context into the deep past and imagined future [9]. Attachment control states do this by bringing together isolated events into a stream of awareness and providing biases which influence all perceptions, memory retrievals and actions. Future work will involve implementing agent-based simulations to produce abstract and simplified versions of the narrative discourse patterns for each AAI classification by augmenting existing attachment models

[24].

REFERENCES

[1] J.C. Ablow, A.K. Marks, S.S. Feldman, and L.C. Huffman, 'Associations between first-time expectant women's representations of attachment and their physiological reactivity to infant cry.', *Child Development*, **84**, 1373–91, (2013). 4.

[2] J.B. Asendorpf and S. Wilpers Church, 'Personality effects on social relationships', *Journal of Personality and Social Psychology*, **74**, 1531–1544, (1998).

[3] A.D. Baddeley, *Working Memory*, Oxford University Press, New York, Oxford, 1986.

[4] L.P. Beaudoin and A. Sloman, 'A study of motive processing and attention', in *Prospects for Artificial Intelligence*, eds., A. Sloman, D. Hogg, G. Humphreys, D. Partridge, and A. Ramsay, 229–238, IOS Press, Amsterdam, (1993).

[5] I. Bretherton and K.A. Munholland, 'Internal working models in attachment relationships', in Handbook of Attachment, (Second edition *, eds. J. Cassidy & P.R. Shaver*, 102–127, Guilford Press, London, (2008).

[6] E.R. Broussard and J. Cassidy, ' Maternal perception of newborns predicts attachment organization in middle adulthood', *Attachment and Human Development*, **12**, 159–172, (2010).

[7] D.M. Buss, 'Social adaptation and the five factors of personality', in *The five-factor model of personality: Theoretical perspectives, ed. J.S. Wiggins*, 180–207, Guilford Press, NY, (1996).

[8] A. Church, 'Culture and personality: Toward an integrated cultural trait psychology', *Journal of Personality*, **68**, 1266–1282, (2000).

[9] M. Donald, *A Mind So Rare: The Evolution of Human Consciousness*, W. W. Norton and Co., 2001.

[10] K.A. Ericsson and W. Kintsch, 'Long-term working memory', *Psychological Review*, **102**, 211–245, (1995).

[11] E. Hesse, 'The adult attachment interview, protocol, method, of analysis, and empirical studies', in Handbook of Attachment, (Second edition *, eds. J. Cassidy & P.R. Shaver*, 552–598, Guilford Press, London, (2008).

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

[13] M. Main, E. Hesse, and N. Kaplan, 'Predictability of attachment behavior and representational processes at 1, 6, and 18 years of age: The berkeley longitudinal study.', in Attachment from Infancy to Adulthood*, eds. K.E. Grossmann, K. Grossmann & E. Waters*, 254–304, Guilford Press, New York, (2005).

[14] D.P. McAdams and J.L. Pals, 'A new big five: Fundamental principles for an integrative science of personality', *American Psychologist*, **61**, 204–217, (2006).

[15] D. Petters, 'Simulating infant-carer relationship dynamics', in *Proc AAAI Spring Symposium 2004: Architectures for Modeling Emotion - Cross-Disciplinary Foundations*, number SS-04-02 in AAAI Technical reports, pp. 114–122, Menlo Park, CA, (2004).

[16] D. Petters, 'Building agents to understand infant attachment behaviour', in *Proceedings of Modelling Natural Action Selection*, eds., J.J. Bryson, T.J. Prescott, and A.K. Seth, 158–165, AISB Press, School of Science and Technology, University of Sussex, Brighton, (2005).

[17] D. Petters, *Designing Agents to Understand Infants*, Ph.D. dissertation, School of Computer Science, The University of Birmingham, 2006. (Available online at http://www.cs.bham.ac.uk/research/cogaff/).

[18] D. Petters, 'Implementing a theory of attachment: A simulation of the strange situation with autonomous agents', in *Proceedings of the Seventh International Conference on Cognitive Modelling*, 226–231, Edizioni Golardiche, Trieste, (2006).

[19] D. Petters, 'Bringing Together Different Pieces to Better Understand Whole Minds', in *From Robots to Humans and Back: Reflections on Hard Problems in the Study of Cognition.*, eds., J.L. Wyatt, D. Petters, and D. Hogg, (2014).

[20] D. Petters, 'Losing Control in the H-CogAff Architecture', in *From Robots to Humans and Back: Reflections on Hard Problems in the Study of Cognition.*, eds., J.L. Wyatt, D. Petters, and D. Hogg, (2014).

[21] D. Petters, 'An encounter between 4e cognition and attachment theory', *Connection Science*, **28**(4), 387–409, (2016).

[22] D. Petters and L. Beaudoin, 'Attachment Modelling: From Observations to scenarios to designs', in *Computational Neurology and Psychiatry: Volume 6 of Springer Series in Bio-/Neuroinformatics.*, eds., P. Erdi, B.S. Bhattacharya, and A. Cochran.

[23] D. Petters and E. Waters, 'A.I., Attachment Theory, and simulating secure base behaviour: Dr. Bowlby meet the Reverend Bayes', in *Proceedings of the International Symposium on 'AI-Inspired Biology', AISB Convention 2010*, 51–58, AISB Press, University of Sussex, Brighton, (2010).

[24] D. Petters and E. Waters, 'Modelling Emotional Attachment: An Integrative Framework for Architectures and Scenarios', in *Proceedings of IJCNN*. IEEE, (2015).

[25] D. Petters, E. Waters, and F. Schönbrodt, '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, (2010).

[26] S. Pinker, *How the Mind Works*, Penguin Books, London, 1998.

[27] A. Sloman, 'The mind as a control system', in *Philosophy and the Cognitive Sciences*, eds., C. Hookway and D. Peterson, 69–110, Cambridge University Press, Cambridge, UK, (1993).

[28] A. Sloman. What sort of control system is able to have a personality?, 1995. (Presented at Workshop on Designing personalities for synthetic actors, Vienna, June 1995).

[29] V. Surendran and L.N. Long, 'Implementing a theory of attachment: A simulation of the strange situation with autonomous agents', in *Proceedings of the 14th International Conference on Cognitive Modelling*, 151–157, Penn State, University Park, PA, (2016).

[30] E. Waters, S. Merrick, D. Treboux, J. Crowell, and L. Albersheim, 'Attachment stability in infancy and early adulthood: A 20-year longitudinal study', *Child Development*, **71**, 684–689, (2000).

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