Automating the Administration and Analysis of Psychiatric Tests: The Case of Attachment in School Age Children

G.Roffo†, **D.B.Vo**†, **A.Sorrentino**†, **M.Rooksby**†, **M.Tayarani**†, **S. Di Folco**‡, **H Minnis**†, **S.Brewster**† **and A.Vinciarelli**†
†University of Glasgow - ‡ University of Edinburgh

ABSTRACT

This article presents the School Attachment Monitor, an interactive system aimed at administering, without the supervision of professional personnel, the Manchester Child Attachment Story Task (a psychiatric test for the assessment of attachment in children). The main goal of the system is to collect, through an interaction process, enough information to allow a human assessor to manually identify the attachment of children. The experiments show that the system successfully performs such a task in 87.5% of the cases (105 of the 120 children involved in the study). In addition, the experiments show that an automatic approach based on deep neural networks can map the information that the system collects, the same that is provided to the human assessors, into the attachment condition of the children. The outcome of the system matches the judgment of the human assessors in 82.8% of the cases (87 of the 105 children for which the system has successfully administered the test). To the best of our knowledge, this is the first time an automated tool has been successful in measuring attachment. This work has significant implications for psychiatry as it allows professionals to assess many more children and direct healthcare resources more accurately and efficiently to improve mental health.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Empirical studies in HCI*;

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300825

KEYWORDS

Attachment, Child Computer Interaction, Mental Health

ACM Reference Format:
G.Roffo†, D.B.Vo†, A.Sorrentino†, M.Rooksby†, M.Tayarani†, S. Di Folco‡,, H Minnis†, S.Brewster† and A.Vinciarelli†. 2019. Automating the Administration and Analysis of Psychiatric Tests: The Case of Attachment in School Age Children. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300825

1 INTRODUCTION

In simple terms, attachment is the psychological construct that accounts for how children perceive the relationship with their caregivers (typically the parents) to be [40]. In particular, the attachment condition of a child is said to be *secure* or *insecure* depending on whether such a relationship is perceived to be beneficial or not, respectively. The main reason why such a distinction is important is that insecure children tend to have a decreased capacity of wellbeing [41] and have difficulties managing their stress responses and coping with trauma, which can lead to aggression and violence [36]. Violent children, upon reaching adolescence and adulthood, have a mortality rate 10 times higher than the rest of the population [33], in part due to increased risk of suicide or violent behavior [28] and in part due to higher incidence of serious health issues such as coronary heart pathologies [8].

The problems above can be addressed, or at least attenuated, if insecure attachment is detected early enough in the life of an individual. However, attachment assessment is an expensive and time-consuming process that requires the participation of highly trained professionals. It is, therefore, performed only on a limited number of children, typically those affected by developmental issues or victims of abuse and neglect. For this reason, this article proposes an interactive system, the *School Attachment Monitor* (SAM), that can automatically administer and analyse, without the supervision of a professional assessor, a test based on the *Manchester Child Attachment Story Task* (MCAST) [13], one of the standard tests most commonly applied to assess the attachment condition of children.

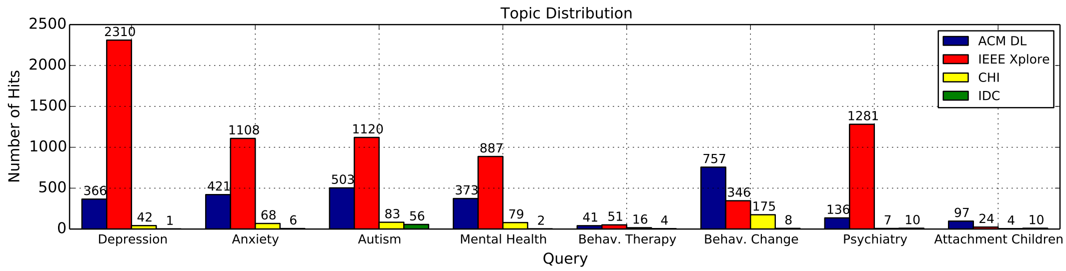

Figure 1: The chart shows the number of hits returned when submitting queries about mental health issues to ACM Digital Library and IEEE Xplore. The chart includes the result for ACM CHI and ACM Interaction Design and Children (IDC).

The main goal of the system is to automatically guide the children through the MCAST procedure while recording enough video data for a professional assessor to identify, at a later stage, the attachment condition. Such an approach replicates the clinical practice of child psychiatrists that first record the test administration with a camera and then analyze the videos to actually perform the assessment.

The main advantage of the system is that it does not require the intervention of a professional assessor during the test administration, a step that typically takes a significant amount of time and cost. In addition, the system analyzes the video data it records and automatically infers the attachment state of the children. The main goal of such a step is to identify the children most likely to be insecure and, hence, more likely to need attention. In other words, the automatic attachment assessment can serve as a triage helping psychiatrists to organize and prioritize their efforts.

The system has been tested by using it to administer MCAST to 120 children (age 5 to 9) randomly recruited among the pupils of several primary schools. The results show it effectively administered the test, meaning that it has collected enough information for an assessor to identify the attachment condition for 105 children, giving a Successful Administration Rate of 87.5%. This is important because it means that the administration efforts of professional assessors can be reduced by 87.5%, thus leaving more time for the actual care of the children.

Furthermore, the automatic analysis approach can correctly infer the attachment state, meaning that it has matched the judgment of the assessors for 87 of the 105 children for which the administration was successful, corresponding to an accuracy of 82.8%. Equally importantly, the approach provides, for each child, a confidence value that estimates how much the outcome of the inference can be trusted. The 68 children corresponding to the highest confidence values have all been classified correctly, meaning that it is possible to establish a confidence threshold above which the accuracy of the approach is 100%. This gives significant benefits to psychiatrists as they can focus their attention on the children with the greatest need of help.

The rest of this article is organised as follows: Section 2 presents an overview of previous work, Section 3 introduces attachment and its assessment, Section 4 presents the system, Section 5 presents experiments and results and Section 6 draws some conclusions.

2 PREVIOUS WORK

During the last decade, the Human-Computer Interaction community has made major efforts towards the development of technologies aimed at dealing with mental health, including "*systems designed for use in prevention of mental illness, standalone computer-based treatment and self-help systems, and systems intended for use in conjunction with face-to-face psychotherapy [...] monitoring of and self-monitoring by clients, communication (such as computed mediated therapy), delivery of content (for example, psycho-educational video material), and interaction with content (for example, interacting with a virtual reality environment to support controlled exposure treatment for phobia)*" [7]. In parallel, computing domains such as Social Signal Processing [37] and Affective Computing [30] have developed approaches for the inference of mental conditions from behavioral data.

As a confirmation of the above, Figure 1 shows the number of hits that *ACM Digital Library* and *IEEE Xplore*, probably the most exhaustive repositories of computing literature, return when submitting queries corresponding to mental health topics. The figure includes the results for the two conferences that appear to be the most relevant to this work, namely the *ACM CHI* and the *ACM Conference on Interaction Design and Children*. The data shows that the number of works mentioning both attachment and children are one order of magnitude less than those mentioning depression, anxiety or autism. This suggests that the problem of attachment in children has attracted limited attention from the community.

Most of the works that the query "*attachment children*" returns try to extend the application of the attachment theory, originally formulated to model the relationship between children and caregivers, to the relationship between users and artifacts [23, 24]. The result is a design theory, the *attachment framework*, aimed at sustainable HCI, i.e., at design and creation of digital artifacts with which users can establish long-term satisfactory use relationships [26, 32]. In such a context, the focus is often on children for two main reasons. The first is that children are more prone to establish affective relationships with objects (toys in particular) [1] and, hence, to transfer their attachment behavior from people to artifacts. The second is that they often exhibit instant satisfaction, the tendency to quickly lose interest in a given item [2].

The possibility of administering a test for the assessment of attachment through an interactive system was proposed earlier [38, 39]. However, these works do not show evaluations or approaches for automatically inferring the attachment of their users. Other works try to support the development of a positive attachment condition through technology [9, 14, 19]. In [9], the core idea is to help the interaction at distance between children and their loved ones via stickers that can be attached to different objects and are equipped to transmit simple stimuli. The children can design both the stickers and the stimuli. According to the authors, this improves the perception of the interaction with others, in particular the parents, one of the crucial aspects in the development of attachment. In a similar vein, the work presented in [19] makes use of mobile technologies to establish better interactions between the members of a family, one of the conditions most likely to lead to secure attachment. The approach proposed in [14] helps parents to tell stories to their deaf children in the same way as they would do with children that can hear. In the intentions of the authors, this can increase the chances of the deaf children developing secure attachment.

Several works address the problem of attachment in the relationships with social robots [15, 16, 18]. The main difference with the approaches inspired by the attachment framework (see above) is that social robots, unlike other digital artifacts, are designed to stimulate the same cognitive and psychological processes that other humans do [5]. Therefore, the attachment behavior observed in human-robot interactions is closer to the one that child psychiatrists observe in their practice. In [15], the focus is on the relationship between the sophistication of a robot and its chances of establishing attachment bonds with its users. The approach proposed in [18] is to build child-robots that can help parents learn how to behave according to the needs of their children and, hence, to develop secure attachment. A similar approach is proposed in [16], where robots displaying different attachment behaviors are used to test how adults react to them.

Finally, only a few works have addressed the automatic inference of the attachment condition from data captured with sensors [27, 35]. In both articles, the experiments show that there is a relationship between ear pulse waves, a physiological measurement related to blood pressure, and attachment. In this respect, the approach proposed in this work (see Section 4) is the first attempt to automatically identify the attachment condition of a child from data, building on standard tools used by psychiatrists.

3 ATTACHMENT AND ITS ASSESSMENT

One of the main tenets of attachment theory is that the "*attachment behavior of different children could be seen as representing different kinds of internal working models of close relationships*" [40]. In other words, the different attachment conditions correspond to different internal models of close relationships that lead the children to behave differently towards their caregivers. In particular, the behavior differences depend on whether the children perceive their caregivers to be consistently and sensitively available or not and, for this reason, the key-step of an attachment assessment process is the inference of the "*internal working models*".

In most cases, the inference of the internal model is performed by asking children to complete story stems that involve attachment relevant situations (e.g., a child does not feel well) to see how the interaction with a caregiver is depicted. Such a methodology, known as *story-stem narrative technique*, has been shown to provide "*a vehicle for accessing the representational world of young children through the developmentally appropriate domain of play, and these techniques have been used successfully in investigations of representation and attachment in both normative and maltreated samples of children*" [34].

The MCAST (*Manchester Child Attachment Story Task*) [13] is one of the story stem narrative techniques that child psychiatrists apply most commonly and it is based on five vignettes portraying children in mildly stressful situations (e.g., a child waking up in the middle of the night after a nightmare). Its administration takes place according to the following protocol:

(1) The psychiatrist shows a vignette to the participant, explains what happens in it and asks to complete the story with the help of two dolls, one representing the child and the other representing a caregiver (typically the mother);
(2) The participant completes the story by representing it with the dolls;
(3) When the participant has finished the story, the psychiatrist asks the participant to say how the child in the story feels;
(4) The participant explains how the child of the story feels;

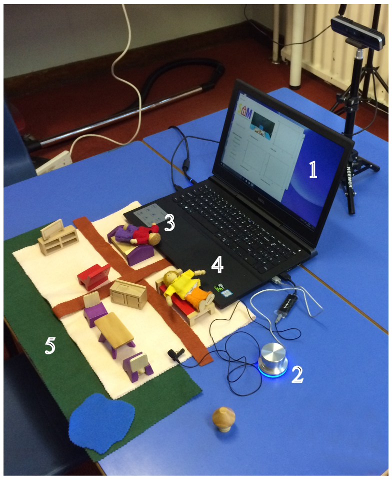

Figure 2: The picture shows the system and how it appears to the children. The computer screen (see element 1 in the picture) displays the videos where actors guide the children through the steps of MCAST, the button (see element 2 in the picture) allows the users to signal that they have completed an MCAST step, the dolls (see elements 3 and 4) and the toy house (see element 5) allow the users to complete the story stems.

(5) When the participant has finished to explain how the child of the story feels, the psychiatrist asks how the caregiver in the story feels;
(6) The participant explains how the caregiver feels;
(7) The psychiatrist goes back to Step 1 and repeats the process for the next vignette until all five vignettes have been addressed.

The use of the dolls allows children to express their feelings and inner working model simply and directly. Psychiatrists record the administration sessions with videocameras so that they can analyze the behavior of the children at a later stage. Furthermore, the video recordings allow one to carefully and reliably analyze the behavior of the children during the administration of the test, one of the main sources of evidence for the inference of the internal working model corresponding to the attachment condition.

Overall, MCAST describes a standard protocol for the administration and classification of attachment status. Specific training and evaluation are required for both of these activities before professionals are certified to MCAST. This is expensive and time consuming. In addition, the analysis of each video takes significant time. These factors mean that there are few trained experts and few children get tested. The Literature shows that MCAST ratings are highly reliable whether conducted via video or computer [25].

4 THE SYSTEM

The main goal of this work is to automate the MCAST administration protocol described in Section 3 with an interactive system capable of guiding the children through the test while capturing enough behavioral information to allow the assessment of their attachment condition. From the point of view of the participants, the system (see Figure 2) includes five main elements, namely a computer screen, a large button to be pushed every time that one of the MCAST stages has been completed, two dolls (child and caregiver), and a toy house equipped with the main pieces of furniture (table, chairs, bed, etc.). In addition, the system is equipped with a camera that captures the upper body of the participants and records their behavior during the administration of the MCAST.

During the administration of the MCAST, the main task of the psychiatrists is to show the vignettes to the participants and to explain what is expected from them. In the case of the automatic administration, the system performs the same task by displaying videos where an actor presents the vignettes and provides the children with instructions on how to proceed. As a consequence, the protocol described in Section 3 becomes as follows:

(1) The participant watches a video where an actor shows a vignette, explains what happens in it and asks to complete the story with the help of two dolls, one representing the child and the other representing a caregiver (typically the mother);
(2) The participant completes the story by representing it with the dolls and pushes a button to signal the system that the task has been completed;
(3) The button activates a video where the actor asks the participant to explain how the child of the story feels;
(4) The participant explains how the child of the story feels and pushes a button to signal the system that the task has been completed;
(5) The button activates a video where an actor asks the participant to explain how the caregiver of the story feels;
(6) The participant explains how the caregiver feels and pushes a button to signal the system the task has been completed;
(7) The button sends the system back to step 1 and repeats the process for the next vignette until all five vignettes have been presented.

In line with the story-stem approach (see Section 3), the system has been designed to appear like a game.

Data Recording and Processing

For the development and testing of the system, the captured video recordings were given to professional assessors for the identification of attachment status using the standard

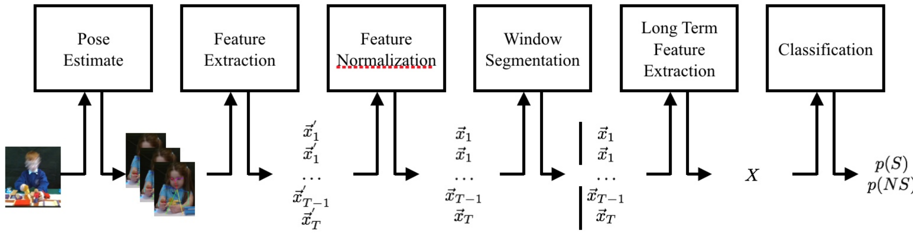

Figure 3: The scheme depicts the video processing approach. A video is fed to OpenPose that generates a sequence of pose estimations. This is transformed into a sequence of feature vectors by the feature extraction step that is followed by the feature normalization. The resulting sequence of normalized feature vectors is segmented into analysis windows and augmented with the long-term features (Long term Feature Extraction). The resulting sequence X of feature vectors is fed to the attachment inference approach (Classification) that gives as output the probabilities $p(S)$ and $p(NS)$ of a child being secure or insecure, respectively.

MCAST protocol. In parallel, they were analyzed automatically by the system. The first step of the automatic process (see Figure 3) is to estimate the pose of the child in each frame of a video. This is done with *OpenPose*, a robust nonparametric algorithm that encodes both position and orientation of human limbs in realtime [6]. In particular, OpenPose takes as input an RGB image and gives as output the coordinates of the joints of any subjects portrayed in it (see Figure 4). After such a step, a video is converted into a sequence of pose estimates represented as t-uples of joint coordinates $\{(x_i, y_i)\}$, $i = 1, \ldots, K$, where K is the total number of joints.

The second step of the processing is the mapping of the pose t-uples above into feature vectors, i.e., into vectors where each component is a physical measurement that accounts for a pose property of interest. In the case of MCAST, the focus is on the position of the hands because the main source of evidence about the attachment condition is the way the participants move the dolls to complete the story stems. For this reason, the features adopted in this work are as follows:

- **Hand positions:** The first four features are the coordinates of the hands (two per hand);
- **Distance between hands:** the fifth feature is the Euclidean distance between the two hands. The main motivation behind the adoption of such a feature is that the physical contact between the dolls (or the lack of it) is one of the main cues that the assessors use to identify the attachment condition;
- **Hands speed:** The sixth and seventh features are the speeds of the two hands. This is an important information because insecure children have been observed to display less motor activation than the others in attachment relevant situations;
- **Hands acceleration:** The eight and ninth features are the accelerations of the two hands. Such features account for the force that the children apply to the dolls and, hence, for their muscular tension, a possible indication of stress and anxiety;
- **Hands 1D-trajectories:** the tenth and eleventh features represent the position (x, y) of the two hands with one feature $t = x + wy$, where w is the width of the image.

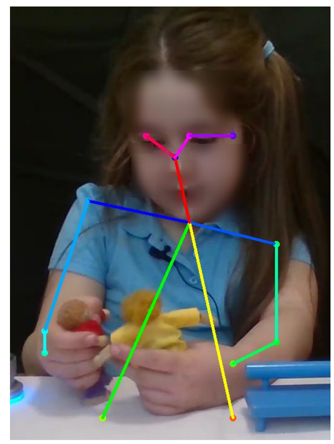

Figure 4: The picture shows how OpenPose processes a generic frame in a video. The dots are the estimated positions of the joints and the lines are the estimated positions of the limbs.

The reason behind the choice of the features above is that non-verbal behaviour, including hand movements, is known to be related to relationship quality and outcome [31].

At the end of the feature extraction every feature x' is normalized as follows:

$$x = \frac{x' - \mu_x}{\sigma_x} \tag{1}$$

where μ_x and σ_x are the average and standard deviation of the feature for a given child. The result is a sequence $X = (\vec{x}_1, \ldots, \vec{x}_T)$ where all features fall in the interval $[-1, 1]$. The normalization is performed to reduce the variance that does not depend on attachment, but on factors such as differences in the position of the camera or height of the child.

A sequence X is the most suitable input for the deep networks used to infer the attachment condition from the videos (see below). However, given that the sequences extracted from a participant tend to be too long for a network to be trained properly, X is segmented into non-overlapping windows of 500 vectors each. In this way, it is possible to analyze the behavior of a child over a sufficiently long term (500 vectors correspond to roughly 16 seconds of video) while avoiding the *vanishing gradient* problem [17], a computational issue that makes the networks ineffective at learning long term patterns when the input sequences are too long (see Section 5 for more details).

In addition to the above, segmenting X into non-overlapping windows allows one to enrich the feature set described above with features that account for longer-term properties:

- **Hands presence:** The twelfth and thirteenth features correspond to the percentage of frames in a window in which the right and left hands are visible, respectively. Such features account for the tendency of the children to use both dolls (both hands visible in a significant fraction of frames) or only one of them (only one hand visible in significant fraction of frames).

Once the two features above have been extracted, every vector of the sequence X includes 13 features expected to capture the way the children play with the dolls.

Attachment Inference

At the end of the feature extraction stage, the original videos have been converted into sequences X of feature vectors. The main reason behind such a choice is that the information conveying the attachment condition of a child is the way she or he moves. Therefore, it is necessary to take into account the temporal evolution of the pose during a video. This suggests that the most suitable models for the inference of the attachment are the *Recurrent Neural Networks* (RNN) and, in particular, the *Long Short Term Memory* (LSTM) networks. These are neural-based computational models capable to capture long-term temporal patterns by taking as input at a given time step t of a sequence their own output at the previous temporal steps $t-1, t-2, \ldots, t-T$.

In particular, given an input sequence $X = (\vec{x}_1, \ldots, \vec{x}_T)$, a RNN estimates a sequence of hidden variables $H = (\vec{h}_1, \ldots, \vec{h}_T)$ and a sequence of output variables $Y = (\vec{y}_1, \ldots, \vec{y}_T)$ by recursively evaluating the following expressions:

$$\vec{h}_t = \sigma(U\vec{x}_t + W\vec{h}_{t-1} + \vec{b}_{\vec{h}}) \tag{2}$$

$$\vec{y}_t = f(V\vec{h}_t + \vec{b}_{\vec{y}}) \tag{3}$$

where the dimension of the vectors $\vec{h}_t$ is the number of neurons in the hidden layers, the dimension of the vectors $\vec{y}_t$ is the number of classes (a sample is assigned to the class corresponding to the highest component), U, V and W are matrices that have the network weights as elements, $\vec{b}$ are the bias vectors, σ is the activation function of the hidden layer (often chosen to be a sigmoid or a hyperbolic tangent), and f is the activation function of the final layer (often a softmax).

The main issue of the RNNs described above is that they suffer from the *vanishing gradient problem* [17], i.e., the tendency of the training (the process through which the network learns to perform a specific task) to become ineffective when the length T of the input sequence increases. However, the literature shows that such an issue can be addressed by introducing *gates* between the input layer and the rest of the network, i.e., processing units that learn how to select useful input information while discarding useless ones. The resulting models are called *Long-Short Term Memory* Networks (LSTMs) [10] and have been used for the experiments of this work (see Section 5).

The information flow inside the network takes place as follows: at a given time step t, vectors $\vec{x}_t$ and $\vec{h}_{t-1}$ are concatenated and given as input to the *forget gate*, the *input gate* and the *output gate* that give as output the vectors $\vec{f}_t$, $\vec{i}_t$ and $\vec{o}_t$, respectively, according to the following expressions:

$$\vec{f}_t = \sigma(W_{xf}\vec{x}_t + W_{hf}\vec{h}_{t-1} + \vec{b}_f) \tag{4}$$

$$\vec{i}_t = \sigma(W_{xi}\vec{x}_t + W_{hi}\vec{h}_{t-1} + \vec{b}_i) \tag{5}$$

$$\vec{o}_t = \sigma(W_{xo}\vec{x}_t + W_{ho}\vec{h}_{t-1} + \vec{b}_o), \tag{6}$$

where σ is a sigmoid, the Ws are weight matrices and the $\vec{b}$s are bias vectors. The vector $\vec{i}_t$ is then multiplied, using the Hadamard product, by a vector $\vec{j}_t = \tanh(W_{xj}\vec{x}_t + W_{hj}\vec{h}_{t-1} + b_j)$, where tanh is the hyperbolic tangent, W is a weight matrix and $\vec{b}$ is the bias vector. The result of such a product is a vector $\vec{c}'_t$.

The three vectors above are used to update $\vec{c}_{t-1}$ and $\vec{h}_{t-1}$ to $\vec{c}_t$ and $\vec{h}_t$, respectively as follows:

$$\vec{c}_t = \vec{f}_t \times \vec{c}_{t-1} + \vec{c}'_t \tag{7}$$

$$\vec{h}_t = \vec{o}_t \times \tanh(\vec{c}_t), \tag{8}$$

where the symbol "×" denotes the Hadamard product. The vectors $\vec{h}$ and $\vec{c}$ capture the temporal information that the sequence X conveys, because the value of $\vec{h}_t$ does not depend only on the input vector $\vec{x}_t$, but also on the vector $\vec{h}_{t-1}$ (the same applies to $\vec{c}_t$). At this point, the process is repeated for the next input vector $\vec{x}_{t+1}$ in the sequence, this time using the vectors $\vec{h}_t$ and $\vec{c}_t$ in equations 4 to 8.

Level	P1	P2	P3	P4
Age	5-6	6-7	7-8	8-9
Total	24 (20.8%)	43 (35.8%)	35 (29.2%)	18 (15.0%)
Female	10 (8.3%)	25 (20.8%)	17 (14.2%)	12 (10.0%)
Male	14 (11.7%)	18 (15.0%)	18 (15.0%)	6 (5.0%)

Table 1: The table shows the age ranges corresponding to the different school levels and the number of participants (with the percentages they correspond to) per level, both in total and split into female and male.

Once the process above has been completed for $\vec{x}_T$ (the last vector in the sequence X), a vector $\vec{h}_T$ is available and it is possible to obtain the output through the following expression: $\vec{y} = g(W_{yh}\vec{h}_T + \vec{b})$, where the function g is a softmax that gives as output a vector in which every component is the probability that the input sequence belongs to a particular class. In the experiments of this work, $\vec{y}$ has two dimensions that correspond to classes *secure* and *insecure*, respectively.

A side product of the process above is that, for every sequence X in the training set, there is a sequence $H = (\vec{h}_1, \dots, \vec{h}_T)$ that captures the possible temporal patterns appearing in the input data. This makes it possible to improve the performance of the LSTMs, by building a stacked network, i.e., by training a second network in which the X sequences are replaced with the H ones while the output vectors $\vec{y}$ remain the same. The information flow is the same as it has been described above and both networks are trained with the same approach (see Section 5).

Overall, this means the network is organized as a stack of LSTMs and, hence, the corresponding model is called *stacked* or *multilayer* LSTM. Such an approach has been introduced in [12] to let a model that is inherently deep in time to benefit from depth in space. Stacking multiple recurrent hidden layers on top of each other [29] allows one to recombine the representations already learned from previous layers to generate representations at higher levels of abstraction.

5 EXPERIMENTS AND RESULTS

The goal of the study presented in this section is to test whether the system described in Section 4 can actually administer the MCAST without the supervision of a professional assessor. In addition, the study has tested whether it is possible to automatically infer the attachment condition of a child from the videos that the system records.

The Participants

The study presented in this work has involved 120 children recruited among the pupils of several primary schools around the city. The third row of Table 1 shows the distribution across the four school levels ($P1$ to $P4$) corresponding to different age ranges. For what concerns the gender, the total number of female and male participants is 64 and 56, respectively (see fourth and fifth row of Table 1). According to a χ^2 test, such a distribution is the same as in the general population and, therefore, the selection process was not biased towards a particular gender. Figure 5 shows the length of the recordings collected for the children involved in the experiments. The average value for the cases in which the material is sufficient for the attachment assessment is 248.2 *sec*, with a standard deviation of 215.4 *sec*.

Effectiveness of Administration

The first question addressed in the study is whether the system described in Section 4 can successfully administer the MCAST without the supervision of a professional. For this reason, the system has been used to administer the test to the sample of 120 children described above. The video recordings collected by the system have been analyzed by a pool of 4 judges that have attended the training course to become professional assessors. At the end of the process, each child has been rated as *secure*, *insecure* or *non-assessable*, where the latter category corresponds to those cases in which the material recorded by the system does not provide sufficient information for a reliable assessment. Whenever there has been disagreement among the judges, consensus has been reached though discussion.

The results show that the data is sufficient to identify the attachment condition in 105 of the 120 cases, corresponding to a *Successful Administration Rate* (SAR) of 87.5%. This suggests that the system can significantly reduce the amount of time required for the administration of the MCAST and, hence, to significantly limit the costs associated to the assessment of attachment. The 15 children for which the assessment was not possible, identified through manual inspection of the data recorded by the system, were unable to use the system (they have not played with the dolls and they have not pushed the button aimed at moving from one step of the process to another).

Table 2 shows the of number of secure and insecure children among the 105 participants above, both in total and for each gender separately. According to a χ^2 test, the distribution across the conditions is the same as in the general population. This confirms that there has not been any bias in terms of attachment condition in the recruitment of the participants. In addition, such an observation suggests that the effectiveness of the system does not depend on the attachment condition of the participants. Finally, Table 3 shows the distribution across the school levels of the 105 children for which the system has successfully administered the MCAST. According to a χ^2 test, the distribution is the same, within a statistical fluctuation, as in the original sample of 120 participants. Therefore, it is possible to say that the effectiveness

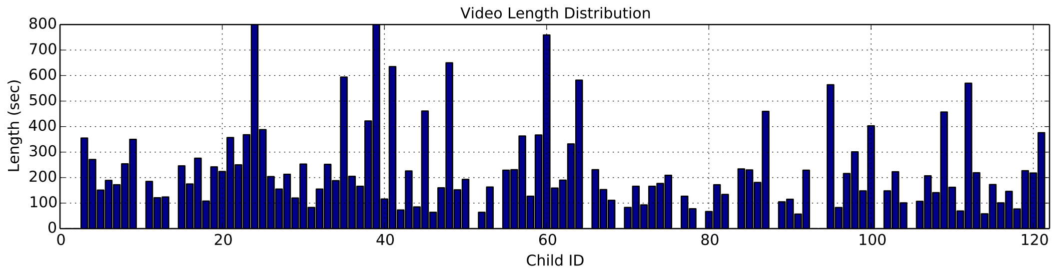

Figure 5: The chart shows the length of the videos that the system has recorded for the children involved in the experiments. When the length is null, it means that the system has not collected information sufficient for attachment assessment. For the sake of clarity, the two bars above $800\ sec$ **(corresponding to** $1,542\ sec$ **and** $1,250\ sec$**) have been cut.**

Attachment	Total	Female	Male
Secure	59 (56.2%)	31 (53.4%)	28 (59.6%)
Insecure	46 (43.8%)	27 (46.6%)	19 (40.4%)

Table 2: The table shows the percentage of secure and insecure children among the 105 **participants for which the system has successfully administered the test, both in total and per gender.**

Level	P1	P2	P3	P4
Number	19	40	30	16
SAR	79.2%	93.0%	85.7%	88.9%

Table 3: The table shows the successful administration rate for children belonging to different school levels.

of the system does not depend on the school level and, indirectly, on the age of the children. This is important because it shows that the system can be adopted over the whole age range the MCAST is designed to cover.

LSTM Training

The second question addressed in this study is whether it is possible to automatically infer the attachment condition from the videos that the system records during the administration of the MCAST. This requires training of the LSTM's and such a task has been performed according to a "*Leave-One-Child-Out*" protocol, i.e., by training the LSTM's over all children except one and then by testing it over the left-out child. Such a process is repeated 105 times and, at each iteration, a different child is left out for testing. In this way, it is possible to test the approach over the whole dataset at disposition while ensuring a rigorous separation between training and test material.

The number of hidden neurons in the LSTM (the dimension of the vectors $\vec{h}_t$ in Section 4) has been set to 128, a value commonly applied in the literature (no attempt has been made to further optimize such a parameter). The training takes place by iteratively changing the value of the network weights (the W matrices and the bias vectors in Section 4) to ensure that the difference is minimized between observed and desired output over the data of the training set. Each iteration of the training process includes two main steps, namely the estimation of the gradient[1] through the back propagation algorithm, and the update of weights and biases, according to the optimization of a predefined criterion[2], using an Adam optimizer [20]. The *learning rate* (a numerical factor that modulates the parameter updates from one iteration to the other) has been set to 2×10^{-4}, a standard value commonly applied in the literature (see. e.g., [3]).

The training is performed with a mini-batch strategy, meaning that the network is trained over small subsets of the training set (the mini-batches) and one iteration of the training process (typically called an *epoch*) corresponds to train the network over all the available mini-batches. The intersection between any two mini-batches is empty while the union of all mini-batches corresponds to the entire training set. The mini-batch approach aims at dealing with the computational issues resulting from the use of large training data sets. In the experiments of this work, every mini-batch includes 800 input sequences and the number of epochs is 30. The mini-batch strategy makes it computationally feasible to train the LSTM's with large amounts of the data at the cost of an acceptable performance loss [21].

Automatic Attachment Inference

Once the LSTM's have been trained, it is actually possible to test the automatic attachment inference. Section 4 shows that the network infers the attachment condition from 500 frames long non-overlapping video intervals (the windows). Therefore, if the video available for one child include N windows, the decision will be made according to a *majority vote*, i.e.,

[1]The derivative of the function implementing the network with respect to weights and biases.

[2]The minimization of cross-entropy between desired and observed output.

α	π	ρ	$F1$
80.2%	67.4%	91.2%	77.5%

Table 4: The table reports the classification performance in terms of accuracy (α), precision (π), recall (ρ) and $F1$ measure.

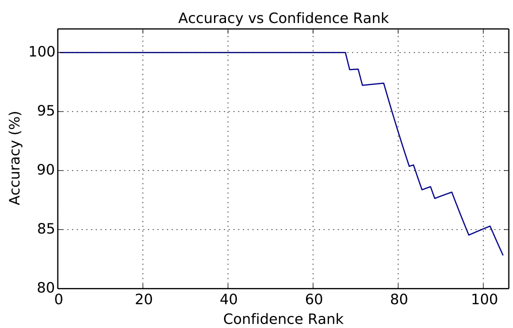

Figure 6: The chart shows the accuracy over the children corresponding to the top r values of the confidence as a function of r (the confidence rank).

using the following equation: $\hat{c} = \arg\max_{c \in C} n(c)$, where $\hat{c}$ is the class the child is assigned to, C is the set of all possible classes (*secure* and *insecure* in this study), and $n(c)$ is the number of windows assigned to class c with $\sum_{c \in C} n(c) = N$. In simple terms, a child is assigned to the class most frequently assigned to the windows extracted from the videos collected during her or his MCAST administration.

The overall classification accuracy, the percentage of times the automatic approach matches the decision of the assessors, is 82.8%, corresponding to 87 of the 105 children for which the identification of the attachment condition was possible. The assessors were present during automatic administration to ensure proper comparison with the AI approach and previous work shows a Cohen kappa 0.67 (considered acceptable) between a previous version of the system and the traditional MCAST [25].

If p_c is the a priori probability of class c (secure or insecure in this study), then a random classifier that assigns a child to class c with probability p_c achieves the accuracy $\hat{\alpha} = \sum_{c \in C} p_c^2$. In the case of this study, $\hat{\alpha}$ is 50.8% and such a value is lower, to a statistically significant extent, than the accuracy achieved with the LSTM's. Therefore, it is possible to say that the automatic assessment approach proposed in this work performs above chance.

Table 4 shows the classification performance of the approach. According to a χ^2 test, there is a statistically significant difference between the accuracy over secure and insecure children, with the former being classified with higher

Level	P1	P2	P3	P4
Accuracy	70.8%	88.4%	91.4%	83.3%

Table 5: The table shows the accuracy over the children belonging to different school levels.

accuracy than the latter. One possible reason is that the number of secure children is higher and, hence, there is more training material for them. Table 5 shows the accuracy for the children belonging to different school levels. According to a χ^2 test, the performance is the same, within a statistical fluctuation, over all levels. Therefore, it is possible to say that the effectiveness of the approach does not depend on the age of the children. This is important because it shows that the system can work with the same effectiveness over the whole age range that the MCAST is expected to cover.

The use of the majority vote allows one to estimate the confidence of the classification approach in its own decision. In particular, if N is the total number of available windows for a child and $\hat{c}$ is the class that maximises $n(c)$, the value $p(\hat{c}) = n(\hat{c})/N$ can be thought of as an estimate of such a confidence. This makes it possible to rank the children according to how certain the classifier is and to check, in correspondence of a given rank r, what is the accuracy of the classifier over the top r ranking children. Figure 6 shows such an accuracy as a function of r.

Overall, Figure 6 shows that the system has an accuracy of 100% over the children that rank in the top 68 positions. Such an experimental observation suggests that the proposed confidence measure, while building upon a baseline approach such as the majority vote, tends actually to be higher when the decision of the approach is correct. In particular, it suggests that whenever the confidence of the system is above the threshold that corresponds to the confidence value observed for the child that ranks 68^{th}, it is possible to say that the approach has made the same decision as the human assessors. However, the plot of Figure 6 cannot exclude that a high confidence value can be associated to a wrong classification outcome. The confidence level is important because it can be used as a criterion to decide whether a case is sufficiently difficult to require expert human assessment. In other words, the confidence value can suggest what are the cases in which the human intervention should have higher priority.

6 DISCUSSION AND CONCLUSIONS

This article shows that an automatic system, the *School Attachment Monitor* (SAM), can successfully administer the MCAST without the supervision of an expert in 87.5% of the 120 cases considered in the experiments. Such a result is important because it shows that the system can potentially reduce by 87.5% the amount of time required to a professional to administer the test and, hence, to substantially reduce the costs associated with the assessment of attachment. In addition, the system can automatically infer the attachment status of a child with an accuracy of 82.8%. Equally importantly, the system accompanies its decision with a confidence level that reliably estimates the chances of the decision being correct (the top 64.8% of the confidence values correspond to correct decisions).

The results above suggest that SAM can substantially reduce the amount of time that professional assessors need to spend in administering the test. Furthermore, the confidence levels can be used to perform a *triage*, i.e., a selection of the cases that, being more ambiguous, require deeper professional attention. The main reason why these results are important is that the early detection of insecure attachment can contribute significantly to improve the life of an individual [8, 28, 33, 41]. Therefore, any means that can help to increase the number of children that, for a given amount of resources, can be screened can have a positive societal impact.

This research is one of the first approaches for automated assessment of attachment. Previous efforts at the crossroad between psychiatry and computing technology have rather focused on depression, anxiety and, in the particular case of children, autism (see Section 2). One possible reason is that the attachment theory was developed around 50 years ago [4], but there was no widespread awareness of the importantance of attachment until recently [22]. Furthermore, it is only in the last decade that some of the main technologies necessary for the automatic analysis of behavior - e.g., pose estimate [6] and deep networks [11] in the particular case of this work - have become sufficiently robust to work outside the laboratory. This is particularly important in the case of the MCAST that, like other story-stem techniques, is expected to look like a game more than like a medical test and, hence, it is expected to take place in an environment that is as friendly and familiar to the children as possible. The data for this work were collected in noisy and dynamic primary schools during term time with many other school activities going on alongside.

An automatically rated measure of child attachment patterns is a significant clinical advance and addresses a gap that had eluded clinicians and researchers for decades. The United Kingdom National Institute for Healtcare Excellence (NICE) Guidelines on Attachment (2015, section 1.3.4) recommend the use of standardised attachment tools (including the Manchester Child Attachment Story Task) for any children who may have attachment difficulties. The fact is that all of the tools recommended are too cumbersome and have training pathways that are too expensive and time consuming for any clinician and, if it were possible for these tools to have been incorporated routinely into clinical practice, this would have happened at least 20 years ago. SAM, therefore, for the first time, offers clinicians a quick and easy measure of child attachment. As children nowadays are used to interacting with technology in the educational setting, as well as in other aspects of their life, such as play and leisure time, they are at ease and motivated in engaging with SAM in clinical assessment.

The capacity and facility to assess attachment difficulties in early life will allow identification of children and their families who require clinician's attention. Decades of research have shown that children at risk for maladaptive attachment are likely to be going through other adverse childhood experiences (ACEs) such as stressors in family functioning or neuro-developmental disorders. Also worth considering here is that effective interventions are available for attachment difficulties, such as video interaction guidance (VIG). This further highlights the value and contribution of SAM, namely for children's social and emotional wellbeing, broader sense of health and quality of life, as well as health economic values such as effective investment of clinician's time and their expertise.

Additional benefits of the use of SAM can also be found in a research setting, as its automated administration ensures the scripts are delivered consistently and with minimum training requirements and costs for the researchers. For these reasons, SAM can be easily included in research protocols, as it can be effectively administered to large-scale samples of children in early middle childhood for screening of their attachment patterns. In addition, the potential for SAM to be used in whole school classrooms, means that it can also be used at a population level as a Public Health Indicator. It will allow children's families and their schools to make more evidence-based decisions on whether a child may benefit from intervention(s) for their apparent problems or concerns. This will give confidence for health and education services to know when to provide reassurance as well as offer help if this is the case.

In conclusion, this article has presented an automatic system that can, for the first time, collect and analyze data for child attachment assessment in psychiatry. The system has matched the judgment of expert human assessors in 82.8% of cases. This work has significant implications for psychiatry as it allows professionals to assess many more children and direct healthcare resources more accurately and efficiently to improve mental health.

ACKNOWLEDGMENTS

This work has been supported by the Engineering and Physical Sciences Research Council (EPSRC) through the grants *School Attachment Monitor* (EP/M025055/1) and *Socially Competent Robots* (EP/N035305/1).

REFERENCES

[1] E.K. Ackermann. 2005. Playthings That Do Things: A Young Kid's "Incredibles"!. In *Proceedings of the ACM International Conference on Interaction Design and Children*. 1–8.

[2] E. Ackermann. 2008. Toys to Fall for or Live with?. In *Proceedings of the ACM International Conference on Interaction Design and Children*. 2:1–2:1.

[3] M. Arjovsky, A. Shah, and Y. Bengio. 2016. Unitary evolution recurrent neural networks. In *Proceedings of the International Conference on Machine Learning*. 1120–1128.

[4] J. Bowlby. 2012. *A secure base*. Routledge.

[5] C.L. Breazeal. 2004. *Designing Sociable Robots*. MIT press.

[6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. 2017. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In *Poceedings of the The IEEE Conference on Computer Vision and Pattern Recognition*.

[7] G. Doherty, D. Coyle, and M. Matthews. 2010. Design and evaluation guidelines for mental health technologies. *Interacting with Computers* 22, 4 (2010), 243–252.

[8] M. Dong, W.H. Giles, V.J. Felitti, S.R. Dube, J.E. Williams, D.P. Chapman, and R.F. Anda. 2004. Insights into causal pathways for ischemic heart disease: adverse childhood experiences study. *Circulation* 110, 13 (2004), 1761–1766.

[9] N. Freed, J. Qi, A. Setapen, C. Breazeal, L. Buechley, and H. Raffle. 2011. Sticking Together: Handcrafting Personalized Communication Interfaces. In *Proceedings of the ACM International Conference on Interaction Design and Children*. 238–241.

[10] F.A. Gers, N.N. Schraudolph, and J. Schmidhuber. 2002. Learning precise timing with LSTM recurrent networks. *Journal of Machine Learning Research* 3, Aug (2002), 115–143.

[11] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. 2016. *Deep Learning*. MIT Press.

[12] A. Graves, A.-R. Mohamed, and G. Hinton. 2013. Speech recognition with deep recurrent neural networks. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 6645–6649.

[13] J. Green, C. Stanley, V. Smith, and R. Goldwyn. 2000. A new method of evaluating attachment representations in young school-age children: The Manchester Child Attachment Story Task. *Attachment & Human Development* 2, 1 (2000), 48–70.

[14] C. Harbig, M. Burton, M. Melkumyan, L. Zhang, and J. Choi. 2011. SignBright: A Storytelling Application to Connect Deaf Children and Hearing Parents. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 977–982.

[15] D.C. Herath, C. Kroos, C. Stevens, and D. Burnham. 2013. Adopt-a-robot: A Story of Attachment. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 135–136.

[16] A. Hiolle, K.A. Bard, and L. Canamero. 2009. Assessing human reactions to different robot attachment profiles. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*. 251–256.

[17] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber. 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[18] H. Ishihara, Y. Yoshikawa, and M. Asada. 2011. Realistic child robot "Affetto" for understanding the caregiver-child attachment relationship that guides the child development. In *Proceedings of the IEEE International Conference on Development and Learning*, Vol. 2. 1–5.

[19] J. Kaye, M. Nelimarkka, R. Kauppinen, S. Vartiainen, and P. Isosomppi. 2011. Mobile Family Interaction: How to Use Mobile Technology to Bring Trust, Safety and Wellbeing into Families. In *Proceedings of the International Conference on Human Computer Interaction with Mobile Devices and Services*. 721–724.

[20] D.P. Kingma and J. Ba. 2014. *Adam: A method for stochastic optimization.* Technical Report. arXiv preprint arXiv:1412.6980.

[21] J. Konečnỳ, J. Liu, P. Richtárik, and M. Takáč. 2016. Mini-batch semi-stochastic gradient descent in the proximal setting. *IEEE Journal of Selected Topics in Signal Processing* 10, 2 (2016), 242–255.

[22] P. Lovenheim. 2018. *The Attachment Effect*. Tarcher Perigee.

[23] A. Meschtscherjakov. 2009. Mobile Attachment: Emotional Attachment Towards Mobile Devices and Services. In *Proceedings of the ACM International Conference on Human-Computer Interaction with Mobile Devices and Services*. 102:1–102:1.

[24] A. Meschtscherjakov, D. Wilfinger, and M. Tscheligi. 2014. Mobile attachment causes and consequences for emotional bonding with mobile phones. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 2317–2326.

[25] H. Minnis, W. Read, B. Connolly, A. Burston, T.-S. Schumm, S. Putter-Lareman, and J. Green. 2010. The computerised Manchester Child Attachment Story Task: A novel medium for assessing attachment patterns. *International Journal of Methods in Psychiatric Research* 19, 4 (2010), 233–242.

[26] W. Odom, J. Pierce, E. Stolterman, and E. Blevis. 2009. Understanding why we preserve some things and discard others in the context of interaction design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1053–1062.

[27] M. Oyama-Higa, J. Tsujino, and M. Tanabiki. 2006. Does a Mother's Attachment to Her Child Affect Biological Information provided by the Child? -Chaos analysis of fingertip pulse waves of children. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vol. 3. 2030–2034.

[28] C.J. Packard, V. Bezlyak, J.S. McLean, G.D. Batty, I. Ford, H. Burns, J. Cavanagh, K.A. Deans, M. Henderson, and A. McGinty. 2011. Early life socioeconomic adversity is associated in adult life with chronic inflammation, carotid atherosclerosis, poorer lung function and decreased cognitive performance: a cross-sectional, population-based study. *BMC Public Health* 11, 1 (2011), 42.

[29] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio. 2013. How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026* (2013).

[30] R. Picard. 1997. *Affective Computing*. MIT Press.

[31] F. Ramseyer and W. Tschacher. 2011. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology* 79, 3 (2011), 284.

[32] C. Remy, S. Gegenbauer, and E.M. Huang. 2015. Bridging the Theory-Practice Gap: Lessons and Challenges of Applying the Attachment Framework for Sustainable HCI Design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. 1305–1314.

[33] S. Scott, M. Knapp, J. Henderson, and B. Maughan. 2001. Financial cost of social exclusion: follow up study of antisocial children into adulthood. *British Medical Journal* 323, 7306 (2001), 191–200.

[34] S.L. Toth, A. Maughan, J.T. Manly, M. Spagnola, and D. Cicchetti. 2002. The relative efficacy of two interventions in altering maltreated preschool children's representational models: Implications for attachment theory. *Development and Psychopathology* 14, 4 (2002), 877–908.

[35] J. Tsujino, M. Oyama-Higa, and M. Tanabiki. 2008. Measurement of ear pulse waves in children: Effect of facing another child and relationship to an action index. In *2008 IEEE International Conference on Systems, Man and Cybernetics*. 2972–2976.

[36] B.A. van der Kolk. 2005. Developmental Trauma Disorder. *Psychiatric Annals* 35, 5 (2005), 401–408.

[37] A. Vinciarelli, M. Pantic, and H. Bourlard. 2009. Social Signal Processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12 (2009), 1743–1759.

[38] D.-B. Vo, M. Rooksby, M. Tayarani, R. Huan, A. Vinciarelli, H. Minnis, and S.A. Brewster. 2017. SAM: The School Attachment Monitor. In *Proceedings of the ACM Conference on Interaction Design and Children*. 671–674.

[39] D.-B. Vo, M. Tayarani, M. Rooksby, R. Huan, A. Vinciarelli, H. Minnis, and S.A. Brewster. 2017. SAM: The School Attachment Monitor. In *Proceedings of the ACM International Conference on Multimodal Interaction*. 497–498.

[40] D. Wilkins, D. Shemmings, and Y. Shemmings. 2015. *Attachment*. Palgrave.

[41] P. Wilson, P. Bradshaw, S. Tipping, G. Der, and H. Minnis. 2013. What predicts persistent early conduct problems? Evidence from the Growing Up in Scotland cohort. *Journal of Epidemiology and Community Health* 67 (2013), 76–80.