Postprint

This is the accepted version of a paper presented at *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on, Sept 21-24, 2015.*.

N.B. When citing this work, cite the original published paper.

Engagement: A Traceable Motivational Concept in Human-Robot Interaction

Karl Drejing*, Serge Thill†and Paul Hemeren‡

School of Informatics

Skövde University

Sweden, Box 408

Email: *karl.drejing@his.se, † serge.thill@his.se, ‡paul.hemeren@his.se

Abstract—**Engagement is essential to meaningful social interaction between humans. Understanding the mechanisms by which we detect engagement of other humans can help us understand how we can build robots that interact socially with humans. However, there is currently a lack of measurable engagement constructs on which to build an artificial system that can reliably support social interaction between humans and robots.**

This paper proposes a definition, based on motivation theories, and outlines a framework to explore the idea that engagement can be seen as specific behaviors and their attached magnitude or intensity. This is done by the use of data from multiple sources such as observer ratings, kinematic data, audio and outcomes of interactions. We use the domain of human-robot interaction in order to illustrate the application of this approach. The framework further suggests a method to gather and aggregate this data. If certain behaviors and their attached intensities co-occur with various levels of judged engagement, then engagement could be assessed by this framework consequently making it accessible to a robotic platform. This framework could improve the social capabilities of interactive agents by adding the ability to notice when and why an agent becomes disengaged, thereby providing the interactive agent with an ability to reengage him or her.

We illustrate and propose validation of our framework with an example from robot-assisted therapy for children with autism spectrum disorder. The framework also represents a general approach that can be applied to other social interactive settings between humans and robots, such as interactions with elderly people.

Keywords—*Human-Robot Interaction; Motivation; Engagement; Multi-modal Data*

I. INTRODUCTION

In everyday interaction, if you feel like you are not attended to, you will likely be uninterested in pursuing further interaction with the person involved [1]. This makes attention one of the critical factors in building interactive agents; but it is not the only prerequisite for meaningful interaction: interaction needs to be maintained. The maintenance, initiation and termination of interaction has been studied in Human-Robot Interaction (HRI) under the concept of *engagement* [2]–[4]. Here, we extend the concept by defining engagement through the lens of motivation theory [5], [6]. We take this perspective because understanding engagement as a motivational process, and how it can be maintained, has had many positive effects in educational settings [7]–[10].

Researchers whose goal is to create social robots use behavioral cues to extract information about an individual's mental states [11]. What is attended? What emotional state is

the individual in? Is the individual enjoying the interaction? We add this question to the list: "can we by noninvasive means measure engagement, seen as a construct derived from motivation theory, so that we can make use of its benefits in interaction between humans and robots?".

Previous research has mainly focused on developing behaviors that initiate interaction (locating a human, turning towards him/her and displaying proper greeting behavior), maintain interaction by tracking the user's face, reformulating requests if the user fails to respond, and terminating the interaction by verbal communication and gestures [3]. Information about that individual's motivational state, which we believe is key to examine engagement, is scarce in this approach. There is a need to regard engagement as a motivational construct to be able to benefit from the positive effects that have previously been found in educational settings, as described below.

We assert that engagement critically includes more than a serial process of initiation, maintenance and termination of collaborative behavior. On the basis of motivational theory, engagement is more appropriately viewed as constructs expressed in terms of behavior type and intensity (fast movements, laughter magnitude, smiles, etc.). We also assert that engagement stems from intrinsically motivated positively affective behavior. Our framework is able to measure these constructs using a noninvasive multi-modal sensory approach, including motion kinematics, eye-tracking, event-segmentation, audio and intervention outcome measurements.

Before presenting the framework in more detail, we describe the positive effects of assessing engagement in interaction and discuss a practical application that can benefit from these positive effects. We then suggest a definition of engagement based on motivation theories. Lastly, we discuss the framework and how data could be gathered to assess its quality.

II. WHY ENGAGEMENT?

Research on engagement, relevant to this context, has been carried out in classroom environments; investigating for instance the positive effects of students being engaged (on-task, interested, etc.), including improved self-efficacy which increases the likelihood of task initiation and decreased anxiety [10], contribution to learning [7] and increased effort, persistence and enjoyment [8], [9]. If a child is anxious, appears uninterested or simply does not enjoy the interaction, then the task will likely fail. The feeling of enjoyment, the urge to put effort into a task and to keep on doing so is more likely to

occur if one can assure that an interaction maintains a high level of engagement. We argue that these effects appear to be overlooked in HRI. Consequently, we propose a method to measure engagement by suggesting a practical application within HRI.

RAT [12] is an technique used in Autism Spectrum Disorder (ASD; [13]) and maintaining engagement is central in this context. One of the challenges of RAT is to create a semi-autonomous robot that can improve social skills in ASD children [12]. Possible benefits include, but are not limited to, a wider range of positive social behavior and increased attention in the interaction sessions (see [14] for a more detailed account). In these therapy sessions, it is desirable to measure the children's emotional, motivational and attentional states so that the robot can act accordingly, reengaging the child if the child disengages. Furthermore, it is desirable that these measurements are done in a noninvasive manner, since people with ASD can be hypersensitive to touch [15]. Currently, noninvasive feature extraction using multi-modal sensory data to assess a user's mental states have shown some promise, but more research needs to be done [16]. We further argue that multi-modal data would contribute to greater redundancy in the framework as well as an opportunity to examine the extent to which engagement can be seen from a broader range of co-occurring behaviors.

As illustrated in the example above, the HRI community has the potential of constructing interactive agents that can assess engagement and utilize its positive power. But to be able to do so, we must first untangle the definition of engagement, which the following section is devoted to.

III. Understanding Engagement in HRI

A. An Approach to Engagement Based on Motivation Theory

When reviewing the literature it is evident that there is "... no single correct- definition of engagement." [7, p. 224]. We seek to make engagement accessible to HRI by suggesting a definition of engagement and how to measure it using data from multiple sources such as the outcomes of an interaction, kinematic data, audio, observer judgments and self-assessments. To achieve this, we need constructs to explore the theoretical foundation that such constructs would be derived from. It is our position that a behavioral definition of *engagement* can be derived from the works of Brehm and Self [5] and Deci and Ryan [6], as the magnitude, or intensity, of the behavior chosen to achieve a goal. More specifically, we propose that: *engagement can be operationalized as the magnitude of an intrinsically motivated behavior that is initiated by an organism to reach a specific goal*. To illustrate this definition, consider the anecdotal account of its discovery: The researcher got an idea which had to be recorded (goal). He then chose to pick up pen and paper to start formalizing the idea by writing (behavior). This behavior was energetic, i.e. he wrote faster, and pressed the pencil harder against the paper than usual (magnitude). In addition, this goal pursuit was enjoyable, brought satisfaction to the discoverer and he did so without external pressure (intrinsically motivated). We would argue that the energetic behavior of this author reflects the level of engagement that was present during this theory-crafting session. Below we explain the theoretical rationale for our proposed definition.

B. Self-Determination Theory

Self-Determination Theory (SDT; [6]) concerns a continuum ranging from amotivation to extrinsic motivation and intrinsic motivation. Each motivational type has some regulatory processes, regulatory styles and a perceived locus of causality. These are factors that promote or hinder internalization of and regulation of behaviors that is exhibited in this motivational continuum [6]. Meyer and Gagne [17] proposed that intrinsically motivated behavior could constitute a framework of engagement, partly because it has been used to guide engagement-relevant variables (in a variety of contexts) such as needs satisfaction, motivational states and psychological and behavioral outcomes. Deci and Ryan [6] also briefly mention that when a subject is intrinsically motivated he/she actively engages in a task that is interesting, promotes growth, is of optimal challenge and is enjoyable. This can be put in contrast to amotivation, where a subject would be forced to commit to a goal where he/she: feels incompetent when it comes to the task to be performed, lacks intention to act, is not expecting a desired outcome and does not value the activity [6]. In this case, the subject would not be likely to act at all or would act without intent [6]. The key is that intrinsically motivated actions, as opposed to externally motivated actions, stems from a need or a "desire" to perform an enjoyable action for its own sake rather than being forced to commit to an action by an outside source. However, this does not imply that an outside source always disrupts intrinsic motivation: Revisiting the example above, a fellow researcher could invite people to a discussion about engagement, which would facilitate the performance of an enjoyable action (discussing the idea).

In the next section we suggest how these theories of motivation can be used to infer the level of engagement from an individual's actions.

C. From Action to Engagement

Skinner, Kindermann and Furrer [18] argue that engagement could be viewed as a quality indicator of the interactions made with the environment. Motivation appears as outward manifestations of engagement, which can be traced and operationalized. Research on motivation underscores the importance of certain psychological qualities of human activity, such as the energy, purpose and durability of that activity [18]. They argue that engagement (together with *disaffection* as its polar opposite) is a central manifestation of an ongoing action. An *action* has certain qualities, such as being directed, sustained and energized. These qualities in turn reflect motivation, together with engagement and disaffection. The concept of *action* in this context does not translate into observable behavior. An action is a super-category to behavior, emotion, attention and goals [7], [19]. Actions are goal-directed, which means that different visible behaviors can belong to the same action; running, spinning and doing jumping jacks are different behaviors belonging to the same action category (getting fit). In addition, the notion that an action is, to various degrees, energized and that this is one of the qualities that reflects motivation, is not a new one. Esqueda (as cited in [5]) found that when the current level of motivation reaches above a certain point, the amplitude of irrelevant responses increase. In that case it was the force participants exerted on a button and the speed of their writing, which is similar to our example above.

A social context can provide socially relevant feedback

which is dependent on the actions that are produced by individuals [20]. Thus, people are able to grasp the intentions of an actor's actions and determine whether the action was on purpose or not. These features allow us to create a motivational construct of engagement, where engagement describes an individual's interaction with the environment, and that engagement should include the initiation and durability (give up/do not give up)of initiated action [7].

It has been demonstrated that we can derive the behavior of an individual by viewing simple kinematic information in point-light displays [21]. Furthermore, such information can be augmented with contextual information from the environment to infer overall goals of an observed action [22]. As such, the ongoing actions described by [19] and [7] above could augment our understanding of engagement by analyzing kinematic data. Data from other sources, such as audio, can be added to increase the redundancy in the framework as well as providing contextual information.

Thus, we suggest that *motivation* provides us with a *goal* as well as modifying the intensity of the *behavior* chosen to achieve a goal. The resulting *action* is goal directed and consists of a behavior that reflects the motivational intensity. This action can then be seen as an *observable behavior* which leads to an *outcome* that is evaluated with respect to the goal. It is from this observable behavior that we can infer the *level of engagement*. It is important to note that it is not only the *intensity* of the behavior that matters, it is also the *behavior in itself*. An action can consist of multiple behaviors, but the motivational state (i.e. the continuum from amotivation via extrinsic to intrinsic motivation) limits what behavior(s) are selected from an organism's repertoire of all possible behaviors. This selection of behavior(s) from a set of all possible behaviors, depending on motivation and goal, has been suggested in neurological studies [23], [24].

This definition serves as a general definition of engagement that extends to a variety of tasks that children perform in therapeutic settings. The framework we propose below is a specific context that we use to test the validity of the definition for joint attention, imitation and turn-taking. A more general definition of engagement should be less context dependent [25], and our future research will examine the degree to which the framework generalizes to other contexts of social interaction.

We have suggested how we can interpret the qualities of observable behavior as levels of engagement. The following section explains the current issues regarding the measurements of engagement from a HRI perspective and introduces a proposal on how to solve these issues using additional sources of data.

D. Issues with Current Measurements from a HRI Perspective

The engagement and motivation literature still relies on observational and/or self-assessment measures [7], [18], [26]–[28]. There appears to be a consensus that the self-assessment measures are a valid way of asserting the level of certain engagement constructs of an individual, i.e., we know if we are motivated, on-task and happy but do not necessarily know why. Some studies have made an effort to clarify whether observations and self-assessment measures correlate. Such a comparison is necessary to ensure that the observer and the observed report similar levels of engagement. An extensive

and ambitious study compared the two and found only weak to moderate correlations [18]. However, as mentioned above, this is a problem from a HRI perspective. To be able to capture when someone becomes disengaged in an interaction, there has to be a better mapping between the observed level of engagement and the level of engagement that is experienced.

There is little to no research on measuring engagement without observers and self-assessment measures. When creating semi-autonomous robots, e.g. for RAT, having an observer on duty to judge engagement would defeat the purpose. One way to solve this issue is to look for co-occurrences between behavior, observers- and self-assessment measurements. Classifying "engagement" behaviors, using motion kinematics, gaze, audio and task outcome data, while finding patterns where observers and self-assessment measures agree would be a first step.

Based on the shortcomings of previous measures of engagement, we need to operationalize the actions mentioned above in [18] into observable behaviors and correlate these to subjective and observers' ratings of engagement levels. We will build on motivation theory to construct a framework, which we propose can extract such data by noninvasive means.

E. Measuring Engagement

To be able to find behaviors and their corresponding intensities that co-occur with engagement we need a ground truth, i.e., an established appraisal of engagement. In clinical settings, and in previous studies on assessing engagement, it is common to use some kind of observer metric [18], [29], [30]. As addressed in the section above, to be able to utilize the benefits of assessing engagement in HRI, we need to work towards metrics based on multi-modal sensory data. This section describes the metric used on to which the observable behaviors will be mapped, in order to move towards a sensor-data only approach to assess engagement in HRI.

Our proposed framework makes use of observers who rate the levels of engagement on a 6-item scale (developed to measure engagement in human-robot interaction with ASD children [29]) and the participants' rate the level of enjoyment they exhibit after each interaction session. The participants' enjoyment ratings will be used as a convergence measure between subjective and observer ratings (see above, [18]).

The observer ratings are then used as a ground truth of the level of engagement at any time during the interaction. Consequently, the multi-modal data is mapped onto the metric to examine what co-occurring behaviors exist for some engagement level as well as the intensity of those behaviors (velocity, acceleration, etc.).

In the next section, we examine in more detail how these are specified in our framework to measure engagement.

IV. FRAMEWORK

Our framework serves as a basis for experimentation to investigate the observable behaviors that can be mapped to different levels of engagement constructs.

The aim of the framework is to:

1) Identify observable behaviors that are engagement-level specific.

TABLE I. THE FRAMEWORK. ON THE LEFT WE SUMMARIZE THE INTERVENTION OUTCOME AND OBSERVABLE BEHAVIORS, FOLLOWED BY THOSE METRICS' VARIABLE VALUES AND UNITS. ON THE RIGHT WE DISPLAY THE OBSERVERS' ASSESSMENTS OF THE ENGAGEMENT LEVEL FOLLOWED BY THAT METRIC'S VALUE. THE ARROW IS TO BE INTERPRETED AS THE MAPPING FROM THE *Measurements* TO THE *Observer Ratings*.

Measurements: Intervention Outcome and Observable Behaviors	Variable Value and Units		Ground Truth: Observer Ratings	Variable Value
Intervention Outcome (ADOS-G score)	0-2			
Gaze (On/Off target)	1/0			
Body (On/Off target)	1/0	$\rightarrow$	Engagement Level	0-5
Arms	m/s, m/s^2			
Hands	m/s, m/s^2			
Emotional Signals	dB, Classification (angry, happy, neutral, sad)			

2) Identify how the parameters of these observable behaviors vary between high and low levels of engagement.

3) Provide a foundation from which we can empirically measure engagement.

Table I. illustrates the core aspects of the framework. The variables in Table I, of which the levels of engagement are mapped onto, are the intervention outcome, self-assessments and multi-modal sensory data. We start by describing how data should be collected and analyzed to validate our framework.

1. Participants might be put in an interaction task with a robot. This stage produces three main categories of data: **A.** The interaction will be recorded so that it can be viewed later by observers, **B.** During the interaction, motion kinematic, gaze and audio data are recorded and, **C.** During the interaction, self-assessments are included to measure participants' enjoyment.

2a. We propose that observers are divided into two groups. The first group is shown recordings of an entire interaction session. Using the 6-item scale, their task is to mark when they feel that an observable behavior match a description for an engagement level (see [21], for a similar segmentation task), thereby creating timestamps for the levels of engagement. The second group is shown recordings from the same interaction, divided into segments based on the segmentation from the first group. The second group then uses the 6-item scale to judge the levels of engagement in these segments. If these two groups ratings achieve a high correlation, then there is a high probability that they have made their ratings based on the information provided in those segments. In addition, this approach also narrows the time frames so that we only need to map the data onto the shorter segments.

2b. Since the framework's starting point regards RAT, we propose that the *Intervention outcome* is measured by the Autism Diagnostic Observation Schedule - Generic (ADOS-G; [31]). This measurement is specific to RAT which implies that other measurements for this variable is to be applied when generalizing the framework to other applications. Using video data, clinicians can assess the interventions by the parameters provided by ADOS-G (described in more detail below). These ratings will provide a measurement of the intervention outcome that serves as one of the predictors of the levels of engagement.

3. Based on the observers' judgments, machine learning algorithms, such as support vector machines or recurrent neural networks, can be used to determine what observable behaviors are more likely to co-occur with specific levels of engagement. Correlations between the participants' self-assessment scores and the observers judgments will also be considered in this approach in order to replicate previous findings [18]. After stage 3 is completed, the validation of the framework will be complete.

Table I. includes the variables that we predict will be qualified indicators for the levels of engagement. This includes possible values (and/or units) a variable can have at any given point during the interaction.

Following this overview of our framework, we present a more detailed account of every step. These are, in order: 1. framework architecture, 2. social interaction tasks (interventions), 3. participant ratings and, 4. observer ratings.

A. Framework Architecture

Our framework is depicted in Table I. where we present our levels of engagement and the inputs that are used to predict them. Following the *Observer Ratings* on the right in Table I, we find 4 major data categories in the *Measurements* column: *Intervention outcome*, which is described in more detail in the *Interventions* section below, *Self-assessment* which are the subjective ratings measured at specific points in time during the interventions (see participant ratings below) and the other data modalities which are either gathered from video or audio.

To summarize the visual and audio modalities: *Gaze* generates a binary input where the target change depending on the state, at some time *t*, in the interaction (the target is defined by what should be attended to at any point in an intervention). *Body* also generates binary input, detecting whether the child's body is facing the correct target. The rationale for choosing these variables is that it provides a reassurance that the remaining data gathered actually concerns the interaction, since it is hard to argue that someone who is not facing the interaction is actually engaged in it.

The *Arm* and *Hand* categories each consists of variables called $\bar{a}_{arm}$ and $\bar{a}_{hand}$ (left side), and $\bar{b}_{arm}$ and $\bar{b}_{hand}$ (right side), respectively. Each of these categories is measured by RGBD-sensors . These measurements are the velocity (m/s), acceleration (m/s^2) and 3D-position of each arm and hand. The rational for choosing these variables is that if we can observe a behavior via video data, we need the intensity of that behavior. By extracting 3D-positions we will be able to map movement patterns formed by the arms and hands and see if there are co-occurrences of patterns with levels of engagement. We will also be able to measure the intensity (velocity and acceleration) of these co-occurring patterns.

Lastly, *Emotional signals* can be extracted from both visual and audio data. The need for the extraction of emotional signals is described above, where we argue that enjoyment is essential for intrinsic motivation which is essential for engagement. It seems plausible to derive a 4 way binary classification of facial features using support vector machine classification [32]. These four states are *angry*, *happy*, *neutral* and *sad*. Classifying emotional states with audio data alone

has proven to be difficult [33], thus audio will be used to provide additional redundancy in the system and provide the 4 way binary classification above with an "intensity" value, i.e. how loud (measured in decibels) the resulting sound of an emotional state is.

The next section concerns a more detailed account of the interaction interventions used in this study.

B. Interventions in RAT

Although our framework can be used as a general tool in assessing engagement in HRI, this section focus on describing the approach in a RAT application. To measure the co-occurrence of behaviors, we propose the use of interventions commonly used in RAT [12]. These interventions consist of an *Imitation*, a *Turn-taking* and a *Joint attention* task.

The *imitation* task consists of movement, gesture and facial imitation (imitate movements, imitate a movement with an object, produce the correct facial expression from a set of emotive cards, etc.); The *Turn-taking* task is conversational and cooperative (e.g. tower of blocks game), and; The *Joint attention* task consists of gaze alteration between an object and a robot. Each task also consists of multiple sub-tasks, meaning that for example the imitation consist of multiple imitation scenarios, not just one.

The ADOS-G tool can be used to quantify the outcomes of the intervention. ADOS-G consists of 4 modules, which are chosen in accordance with the age of the child. These modules contain items assessed by a clinician. There are 28-31 items (depending on the module) which include assessments of e.g. anxiety, initiation of joint attention and unusual eye contact. These items are measured on a 3 point scale (from 0 to 2) where 0 indicates no autism related abnormality and 2 indicates definite evidence for such abnormality. The total ADOS-G score will serve as predictors for the levels of engagement.

1) Participant Ratings: After each intervention, the children conduct a self-report on how much fun they had during the intervention. Working with children, one must consider that they live in their own world, which requires special adaptations for the researcher [34]. To this end, [35] designed a tool to measure "emotional engagement" and concentration (smiles, laughter, fingers in mouth, etc) with a "Fun-o-meter". The one we adopt here is a an ordinal 5-level smiley scale, with a sad face and a happy face respectively at the corresponding ends of the scale. Note that this should not serve as a direct measurement of the perceived engagement. Instead, this measure should be considered a failsafe in that it should not diverge with the observers' ratings of enjoyment.

2) Observer Ratings: As stated previously, observers will conduct an event segmentation task in order to find co-occurrences between levels of engagement and behavior. Their task is to use a scale developed for measuring engagement of ASD children during human-robot interaction [29]. As shown in Table II. this is a 6-item scale that puts emphasis on the level of compliance. To revisit our definition of engagement, looking at level 4 and 5 of the scale, we can see that they involve a willingness to interact with the robot. Either this is done in a level 5 manner (spontaneously, which would be purely intrinsically driven) or in level 4 manner (interaction after experimenters request, showing a somewhat extrinsic to weak intrinsic drive). The other levels suggest that motivation becomes more and more extrinsically driven or that it fades

TABLE II. LEVELS OF ENGAGEMENT

Rating	Meaning	Description
0	Intense noncompliance	The child walked away from the place in which the robot/human interaction took place.
1	Noncompliance	The child refused to comply with the experimenters request to play with the robot/adult
2	Neutral	The child complied with instructions to play the game with the robot/adult after several prompts from the experimenter.
3	Slight interest	The child required two or three prompts from the experimenter before responding to the robot/adult.
4	Engagement	The child complied immediately following the experimenters request to play with the robot/adult.
5	Intense engagement	The child spontaneously engaged with the robot/adult.

completely in which case the child has lost all its interest in the interaction. This scale will be used as an annotation tool by the observer. Using annotation programs, such as ELAN[1], observers in group one are to be shown recordings of an interaction with a child. Their task is to assign the level and duration of engagement throughout the whole interaction. The second group will then be shown an edited version of the interactions based on the ratings of the first group. The editing will be made based on when one level of engagement transitions to another (5 seconds before and after a transition). These ratings will then be judged for inter-rater reliability which, if high enough, will ensure us that the rater based their judgments of those engagement levels on the data available in the 10 second edited video.

V. CONCLUSIONS: THEORETICAL CONTRIBUTIONS AND FUTURE WORK

The main contributions of our proposed framework and method are twofold: First, it suggests how engagement can be interpreted as the intensity and selection of behaviors, that can be measured via noninvasive means. This definition is contrasted with definitions made in previous research; Instead of claiming that engagement is built on constructs such as attention, pride, anxiety etc. [18], [30], we suggest the use of motivation theory and adding observable behaviors with their corresponding traceable intensities to measure engagement in HRI. Engagement is often a matter of degree, and the ability to measure this degree should be a part of any artificial system that responds to engagement in social interaction.

Second, it could provide us with a database of behaviors that are more likely to occur with higher levels of engagement. Such a database could prove valuable to further examine if these behaviors apply to other populations, such as normally developed children of the same age.

As shown above, engagement is not a readily graspable subject: Engagement could be comprised of behaviors that together form a complex system that we need to understand in order to avoid disengagement.

This paper suggest a definition of engagement based on motivation theory, as well as providing a framework to measure engagement as the intensity and type of goal directed behavior. We also need to know when a human disengages,

---

[1]https://tla.mpi.nl/tools/tla-tools/elan/

and how to reengage that human. Our current setup provides the opportunity to answer the question of what cues are present when a human disengages. In addition it also gives us the opportunity to examine the behaviors that signal a drop in engagement. Thus there is a possibility to refine the dynamics of future interaction as well; if a human is attending and is initiating actions, but shows signs of frustration or anxiety, the robot has more data to use to adapt its behavior than if we only had data based on whether the human is attending to that interaction or not. This framework will ultimately improve the interaction between humans and robots by including more dimensions, i.e. behaviors, to engagement than are currently being considered.

REFERENCES

[1] S. Baron-Cohen, *Joint attention: Its origins and role in development.* Hillsdale, NJ: Erlbaum., 1995, ch. The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology., p. 4159.

[2] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on.* IEEE, 2010, pp. 375–382.

[3] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

[4] H. Salam and M. Chetouani, "A multi-level context-based modeling of engagement in human-robot interaction," *International Workshop on Context Based Affect Recognition.*, 2015.

[5] J. W. Brehm and E. A. Self, "The intensity of motivation," *Annual review of psychology*, vol. 40, pp. 109–131, 1989.

[6] E. L. Deci and R. M. Ryan, "The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior," *Psychological inquiry*, vol. 11, no. 4, pp. 227–268, 2000.

[7] E. Skinner, A. Kindermann, P. Connell, and G. Wellborn, "Engagement and disaffection as organizational constructs in the dynamics of motivational development," *Handbook of motivation at school*, pp. 223–245, 2009.

[8] J. L. Meece, E. M. Anderman, and L. H. Anderman, "Classroom goal structure, student motivation, and academic achievement," *Annu. Rev. Psychol.*, vol. 57, pp. 487–503, 2006.

[9] J. Elliot, A., "A conceptual history of the achievement goal construct," in *Handbook of competence and motivation*, J. Elliot, A. and S. Dweck, C., Eds. New York: New York: Guilford, 2005.

[10] H. Schunk, D. and H. Pajares, "Competence perceptions and academic functioning," in *Handbook of competence and motivation*, J. Elliot, A. and S. Dweck, C., Eds. New York: New York: Guilford, 2005.

[11] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction.* ACM, 2006, pp. 33–40.

[12] S. Thill, C. Pop, T. Belpaeme, T. Ziemke, and B. Vanderborght, "Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook," *Paladyn*, vol. 3, no. 4, pp. 209–217, 2012.

[13] B. Scassellati, H. Admoni, and M. Matarić, "Robots for use in autism research," *Annual Review of Biomedical Engineering*, vol. 14, no. 1, pp. 275–294, 2012.

[14] B. Vanderborght, R. Simut, J. Saldien, C. Pop, A. S. Rusu, S. Pintea, D. Lefeber, and D. O. David, "Using the social robot probo as a social story telling agent for children with asd," *Interaction Studies*, vol. 13, no. 3, pp. 348–372, 2012.

[15] S. Tomcheck and W. Dunn, "Sensory processing in children with and witout autism: A comparative study using the short sensory profile." *American Journal of Occu*, vol. 61, no. 2, pp. 190–200d, 2007.

[16] D. Vaufreydaz, W. Johal, and C. Combe, "Starting engagement detection towards a companion robot using multimodal features," *Robotics and Autonomous Systems*, 2015.

[17] J. P. Meyer and M. Gagne, "Employee engagement from a self-determination theory perspective," *Industrial and Organizational Psychology*, vol. 1, no. 1, pp. 60–62, 2008.

[18] E. Skinner, A. Kindermann, T., and J. Furrer, C., "A motivational perspective on engagement and disaffection: Conceptualization and assessment of childrens behavioral and emotional participation in academic activities in the classroom," *Educational and Psychological Measurement*, 2009.

[19] E. E. Boesch, *Psychopathologie des Alltags.* Huber, 1976.

[20] J. Brandtstädter, "Action perspectives on human development," *Handbook of child psychology*, 1998.

[21] P. E. Hemeren and S. Thill, "Deriving motor primitives through action segmentation," *Embodied and grounded cognition*, p. 63, 2010.

[22] S. Thill, H. Svensson, and T. Ziemke, "Modeling the development of goal-specificity in mirror neurons," *Cognitive computation*, vol. 3, no. 4, pp. 525–538, 2011.

[23] C. B. Holroyd and N. Yeung, "Motivation of extended behaviors by anterior cingulate cortex," *Trends in cognitive sciences*, vol. 16, no. 2, pp. 122–128, 2012.

[24] F. Kouneiher, S. Charron, and E. Koechlin, "Motivation and cognitive control in the human prefrontal cortex," *Nature neuroscience*, vol. 12, no. 7, pp. 939–945, 2009.

[25] C. L. Sidner, C. Lee, and N. Lesh, "Engagement rules for human-robot collaborative interactions," in *IEEE International Conference On Systems Man And Cybernetics*, vol. 4, 2003, pp. 3957–3962.

[26] T. Wang, M. and A. Fredricks, J., "The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence," *Child Development*, vol. 85, pp. 722–732, 2014.

[27] P. Blatchford, P. Bassett, and P. Brown, "Examining the effect of class size on classroom engagement and teacherpupil interaction: Differences in relation to pupil prior attainment and primary vs. secondary schools," *Learning and Instruction*, vol. 21, no. 6, pp. 715 – 730, 2011.

[28] W. Coster, G. Bedell, M. Law, M. A. Khetani, R. Teplicky, K. Liljenquist, K. Gleason, and Y.-C. Kao, "Psychometric evaluation of the participation and environment measure for children and youth," *Developmental Medicine and Child Neurology*, vol. 53, pp. 1030–1037, 2011.

[29] E. S. Kim, R. Paul, F. Shic, and B. Scassellati, "Bridging the research gap: Making hri useful to individuals with autism," *Journal of Human-Robot Interaction*, vol. 1, no. 1, 2012.

[30] A. Fredricks, J., "School engagement: Potential of the concept, state of the evidence," *Review of Educational Research*, vol. 74, pp. 59–109, 2004.

[31] C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedulegeneric: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[32] E. Mower, M. J. Matarić, and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1057–1070, 2011.

[33] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 1, pp. 39–58, 2009.

[34] A. Druin, "The role of children in the design of new technology," *Behaviour and information technology*, vol. 21, no. 1, pp. 1–25, 2002.

[35] J. Read, S. MacFarlane, and C. Casey, "Endurability, engagement and expectations: Measuring children's fun," in *Interaction design and children*, vol. 2. Shaker Publishing Eindhoven, 2002, pp. 1–23.