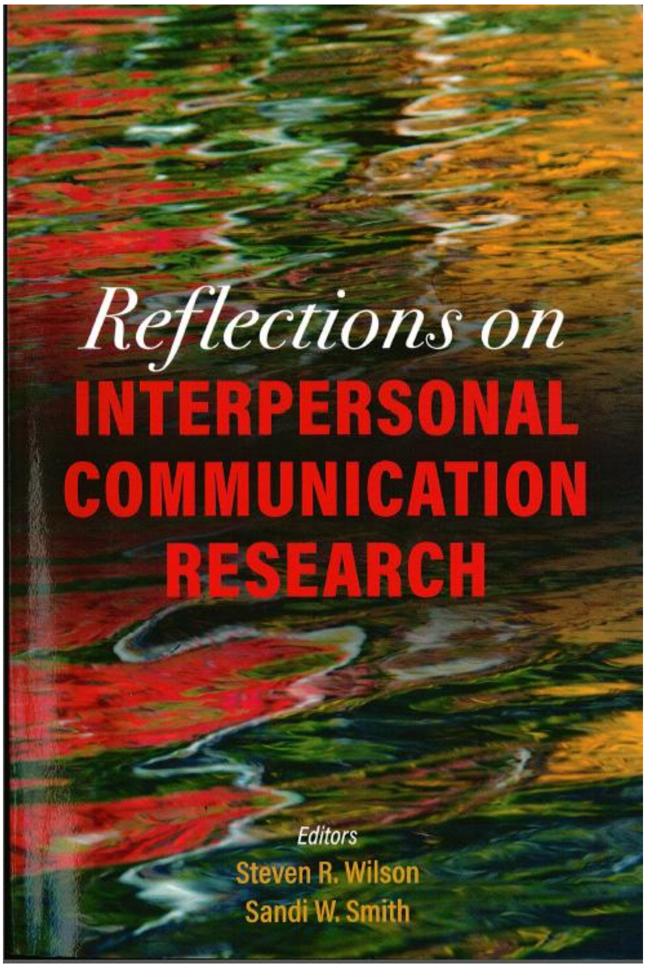

Reflections on Interpersonal Communication

S.W. Wilson and S.W. Smith

CITE AS:

Bente, G. (2019). New tools – new insights: Using emergent technologies in nonverbal communication research. In S.W. Wilson and S.W. Smith (Eds.), Reflections on Interpersonal Communication (pp. 161-188). San Diego: Cognella.

New Tools – New Insights.
Using Emergent Technologies in Nonverbal Communication Research.

Gary Bente

Humans are social animals (Aaronson, 1972). We live in groups and our survival and achievements very much depend on our ability to coordinate our actions with others and collaborate towards shared goals. To this end, evolution has equipped us with unique cognitive capacities and sophisticated means of communication which allow us to understand each other and to jointly adapt to complex affordances in our environment. Beyond language, which sets us apart from other species, our interpersonal relations still largely depend on the subtle, yet powerful signals of our bodies and faces. The impressions of others that we form on basis of their physical appearance and nonverbal behavior are extremely fast and persistent and they often impose on us as facts with seemingly irrefutable certainty. As Asch (1948) put it seven decades ago:

> *"We look at a person and immediately a certain impression of his character forms itself in us. A glance, a few spoken words are sufficient to tell us a story about a highly complex matter. We know that such impressions form with remarkable rapidity and great ease. Subsequent observations may enrich or upset our first view, but we can no more prevent its rapid growth than we can avoid perceiving a given object or hearing a melody"(p. 258).*

When making inferences about others' trustworthiness or kindness, or inferring their intentions or emotions, the impact of the visuals is so immediate and strong that we believe to literally "see" the others' state of mind directly without any intermediate cognitive effort or conscious control. The immediacy and automaticity on the receiver side constitute some risks and opportunities on the sender side as they determine others' perceptions of ourselves and offer a powerful means to influence their attitudes and behaviors towards us (Patterson, 1990). Argyle (1975) characterized the reason for the prominent role of nonverbal signals in interpersonal communication:

> *"While words can express attitudes to others, bodily signals have certain clear advantages; first they are stronger, and have a more immediate impact; secondly, negative signals can be used outside full conscious awareness; thirdly, signals negotiating relationships can be used subtlety, again outside awareness, and can be easily withdrawn" (S, 132).*

Its powerful effects and the subtle, complex, and often puzzling dynamics certainly explain the persistent research interest in nonverbal communication. Yet, they are also the major reason for the slow progress in unravelling its secrets, which I encountered when I entered the field in the early 80s (Cappella & Palmer, 1990; Donaghy, 1989). My research program during over almost four decades has aimed to provide solutions to this problem. It has been guided by three major questions: How can we measure the physical properties of visible behavior independent from its socio-emotional effects? How can we then identify the cues and features that drive our perceptions and interpersonal responses? And, how can we exert experimental control over specific behavioral aspects embedded in a complex multichannel activity in order to isolate their effects? In this chapter, I outline the basic methodological challenges I have encountered in this endeavor and show how emergent technologies, such as character animation, avatars, and motion capture, can help to solve problems in measurement, analysis, and experimental control.

The Methodological Challenge: Separating Description and Evaluation

My research in nonverbal communication started with my dissertation project in clinical psychology on social influence processes and nonverbal dynamics in therapist-patient interactions (Bente, Frey & Hirsbrunner, 1984). At that time, emergent video technology had already improved the starting point of nonverbal communication research substantially. It allowed us to revisit the complex stream of behavior multiple times and to spot even the most subtle and fugitive nonverbal phenomena. Yet, I had to learn the hard way that the repeated inspection of video recordings did not per se unravel the secrets of nonverbal communication. Video is not data and it left unresolved the more fundamental methodological problem of transliteration of the wealth of observable details and interdependencies into analyzable data protocols. Various coding strategies had been suggested at that time to capture the details of nonverbal communication, which all left me dissatisfied with regard to their resolution and accuracy (cf. Frey, Hirsbrunner, & Bieri-Florin, 1979). Figure 1 systematizes the most prominent coding approaches and the respective information loss they accept as well as possible measurement alternatives. As will show, the most important distinction to be made here is the one between descriptive and evaluative procedures. I use the term "transcription," and synonymously, "(physicalistic) notation," in contrast to "coding" to classify alternative measurement procedures, which rely on human observers, but are widely unaffected by implicit or explicit assumptions about the meaning of the observed nonverbal behaviors.

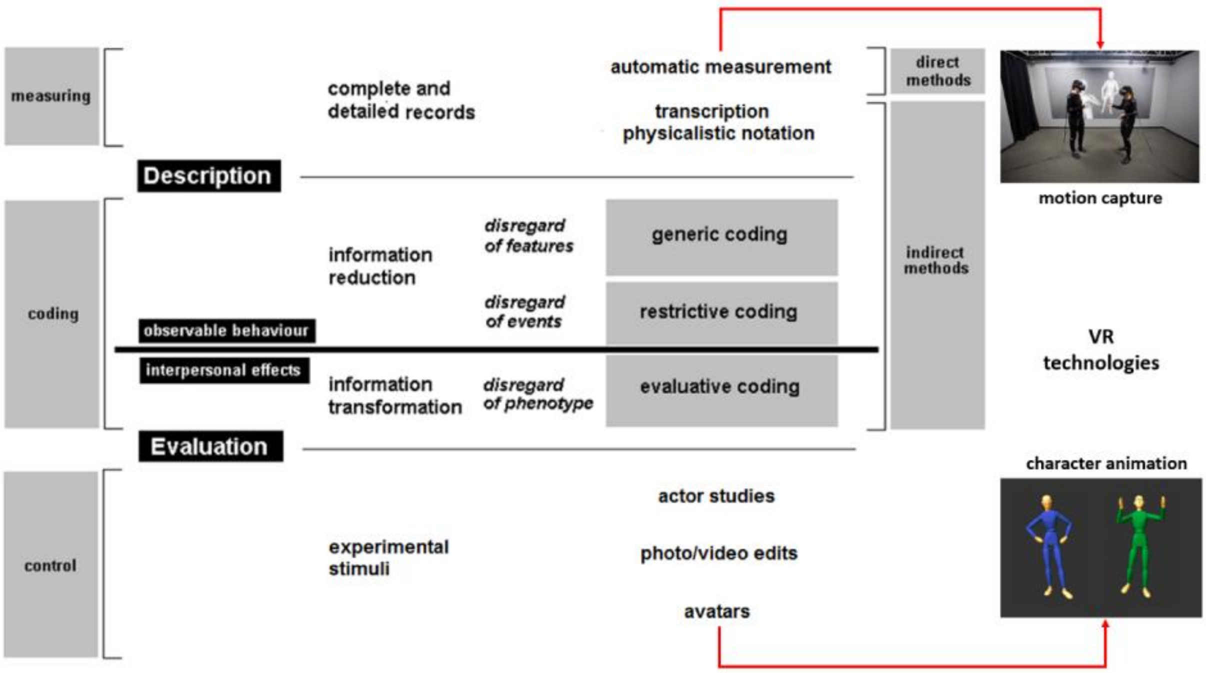

Figure 1: Taxonomy of measurement approaches and experimental control techniques in nonverbal communication research (explanation in the text).

Generic coding refers to the use of broad behavioral categories that are so extensive that they include a variety of clearly distinguishable behaviors. For instance, coding behavior such as kneeling, sitting, or standing during a religious ceremony might be an easy task, but the data would not contain any information about individual postures and movements. Information reduction in generic coding occurs through a disregard of behavioral features. Galliker (1990) characterized generic coding as a "superficial approach, which is indifferent to the complexity of the object" (p. 228).

Restrictive coding refers to the notation of a few, yet precisely defined nonverbal behaviors, such as an eye blink, a head nod, or a smile. Information reduction concerns the disregard of events and their interdependencies. For instance, a smile can mean different things when combined with eye contact than with an averted gaze. Donaghy (1989) concludes:

> *"Restrictive coding is adequate when only very specific types of nonverbal communication are of interest, but this means neglecting the interrelationship that all nonverbal behaviors have to one another" (p. 299).*

Overall, generic and restrictive coding subscribe to a descriptive standard. However, they embrace the possibility of losing information on the way from video to data. Selection or abstraction require explicit coding rules based on theoretical presuppositions, or ideally on factual knowledge of the relevant behavioral elements, which is de facto often not given (Cappella & Palmer, 1990).

In stark contrast to generic and restrictive coding, *evaluative coding* completely skips the descriptive phase of research and transforms observables into psychological categories (see Figure 1). Evaluative coding requires the coder to perform a subjective analysis of the observed behavior and to score its perceived appeal (e.g., in terms of activity, dominance, affiliation, aggression, etc.). Frey, Hirsbrunner, and Bieri-Florin (1979) posit that evaluative coding methods

> *"… provide information about how the behavior 'affects' the observer, but they do not give any information on the question of which concrete non-verbal patterns of behavior produced this impression" (p. 198).*

In this sense, evaluative coding does not represent a measure of nonverbal behavior but a measure of its interpersonal effects. The collection of evaluative data, thus, defines a separate research problem, which requires different solutions than the description problem. Causal inferences from the physical properties of behavior to the psychological effects require precise knowledge of the stimulus components and ideally rigorous experimental control over the cues under inspection (cf. Berry, 1990). This rigorous control, however, poses serious problems. On the one hand, the common strategy to use actors to manipulate specific details of behavior has not proven functional in this regard. For instance, Lewis, Derlega, Shankar, Cochard, and Finkel (1997) found in that instructed actors were barely able to focus their behavioral variation intentionally on one single aspect of nonverbal behavior.

On the other hand, the predominant use of video or photo stimuli has been burdened with a serious confound. Video recordings and photos do not only portray the nonverbal behavior, they also reveal person characteristics, such as gender, race, culture, age or attractiveness, which can be expected to activate stereotypes and evaluation biases (cf. Puccinelli, Tickle-Degnen, & Rosenthal, 2003). Such biases can override the effects of the observed behavior, for instance when judging the aggressiveness of male and female targets.

Summing up, what might appear trivial to the reader came to me as an insight, namely that I had to make an essential distinction between description and evaluation in my research efforts. The description states in physical terms which observable behavior occurs. The evaluation determines in a second step (independent from the first step) what this behavior means, or what interpersonal effects emanate from it. In this sense, both approaches to communication are equally relevant, yet they represent different phases in the research process and have to be strictly separated in the measurement operation to avoid confounds and circularities (Galliker, 1990). Moreover, both tasks have posed specific methodological problems, on the one hand, asking for solutions to objectively measure the physical properties of nonverbal behavior, and on the other hand, to control specific aspects of nonverbal communication experimentally avoiding any confounds.

Addressing the Description Problem

Time Series Notation: A First Descriptive Account of the Physics of Behavior

As shown in Figure 1, automatic measurement and physicalistic notation are suggested as solutions to the description problem. It is important to note that both methods adhere to the same representational principles, which imply the objective, exhaustive, and detailed documentation of the physical features of movement behavior independent of any presuppositions of relevance and meaning. The major difference may only be that notation still needs a human observer. Given the lack of automatic measurement devices in those days, I started my work using the “Bernese Notation System” for human movement behavior, which in my view perfectly matched the criteria for an objective and comprehensive description method (Frey, Hirsbrunner, Florin, Daw, & Crawford, 1983). The notation principle of the Bernese System is based on the decomposition of movement into its temporal and spatial components. Simply spoken, movement is transcribed from video as a sequence of positional states of the various body parts. The positional states are represented as so-called flexion levels, which are analogous to rotation angles around the three axes of a global coordinate system. Detailed descriptions of the notation dimensions and procedures can be found in Frey, et al. (1983; see also Bente, Donaghy & Suwelack, 1998).

The first study I conducted using the Bernese System was my dissertation project (Bente, Frey & Hirsbrunner, 1984). It was a single case process study comparing an experienced and an inexperienced Rogerian client-centered therapist. Movement transcription was done for the first, middle, and last three minutes of each of the first three therapy sessions. The overall 54 minutes took about 150 hours of transcription. While this was shocking for many of my colleagues, I was quite happy to look at a unique data set after only about one month of work. The results of the study changed my thinking about the non-directive nature of the therapy approach under investigation and underpinned the paramount importance of nonverbal behavior in the evolving therapist-patient relationship. Surprisingly, the experienced therapist used a variety of verbal influence strategies including empathetic verbalizations as well as coercive interventions with explicit criticism. While none of the influence attempts changed the client’s verbal self-disclosure, the coercive therapist behavior had a massive influence on the client’s movement activity, which could be interpreted as high arousal and uncertainty as shown in Figure 2 (cf. Bente, Frey, & Hirsbrunner, 1984). Although limited in its generalizability, the data quality of the single case study was encouraging. It showed that human movement carries essential information about the relational dynamics in dyadic interactions and that time series notation offered unprecedented opportunities to study these dynamics.

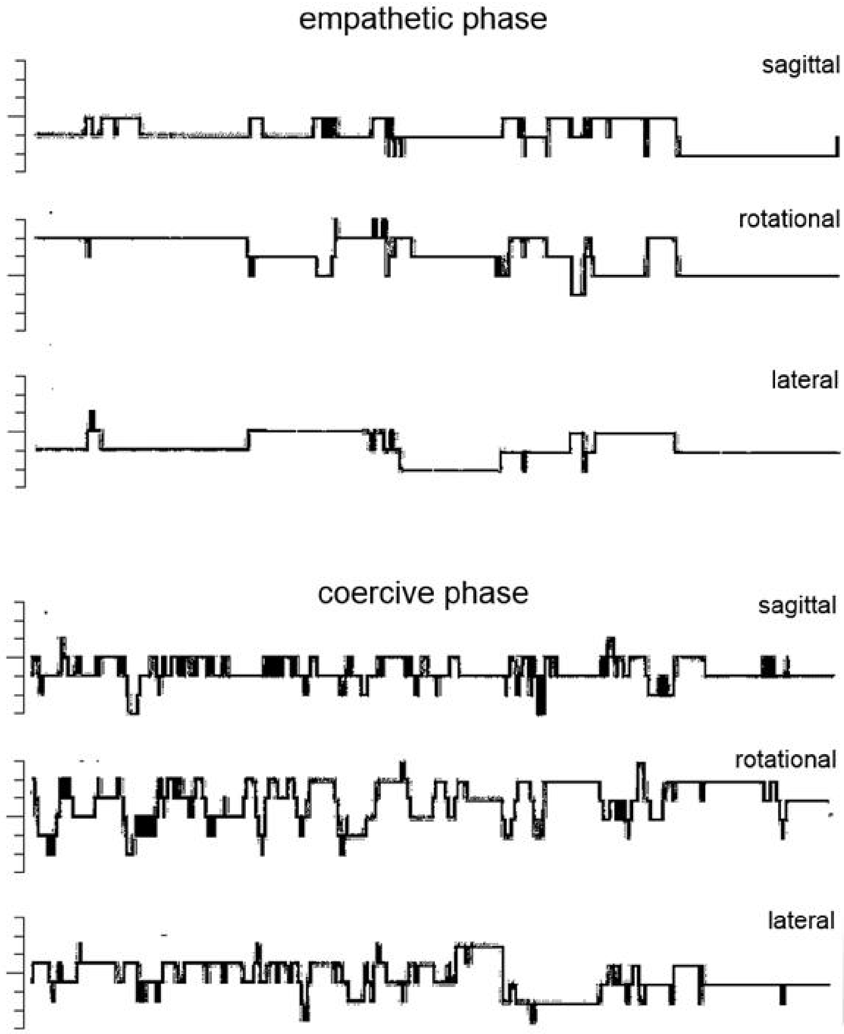

Figure 2: Patient's three-dimensional head movement activity during phases with empathetic and coercive therapist behavior. (sagittal: raise/lower; rotational: right/left turn; lateral: right/left tilt). The X-axis represents the time (3 minutes). The Y-axis represents the flexion levels with the zero point in the three dimensions defining an upright, unrotated and untilted head position with regard to the frontal video camera viewpoint.

Importantly, the quality of the descriptive data triggered a simple idea. If physicalistic notation generated precise information about translations and rotations of the body parts, it should be possible to (re-)translate this data into the visible reality of moving and interacting virtual human bodies (cf. Bente, 1989). Such a simulation of behavior appeared to me as one of the biggest challenges for science in general and the ultimate check for the reliability and validity of our measures (cf. Hut & Sussman, 1987). In my case, this meant using translation and rotation data to drive computer animations of human body models. If this would be possible in real-time, this could enable us to instantly check the correctness of our data entries by comparing the result of the animation with the video. There would even be far-reaching implications for the possibility of stimulus control in experimental effect studies.

With a Little Help from the Avatar: Computer-Assisted Transcription

First trials to use the data of the Bernese Notation System for computer animations of human body models were promising (Bente, 1989). I used a simple homemade wire frame model of a human head and a torso animated by a fast (for the era) algorithm programmed in assembler under MS-DOS on a 4.77 MHz IBM PC with Intel 8088. The program was conceptualized as a spreadsheet to facilitate time series notation. Data entries were converted to Euler angles in real-time and the human 3D model was rotated accordingly allowing me to compare its new position to the video. The completed sequence then could be run as a continuous animation, providing a further possibility to check movement realism. One could say it was an early version of a “virtual social reality” component using an avatar for coder feedback and for experimental stimulus control. Figure 3 shows a scene shot of an animation sequence displaying the head movement of a male and female interactant overlaid to a static image of the bodies. As can be seen the 3D model of the head was extremely simple, but seeing it in motion was impressive. Although the notation procedure still took as long as before, the proof of principle was convincing and advancements in technology could be expected to allow for higher avatar realism, faster animation algorithms and better interfaces.

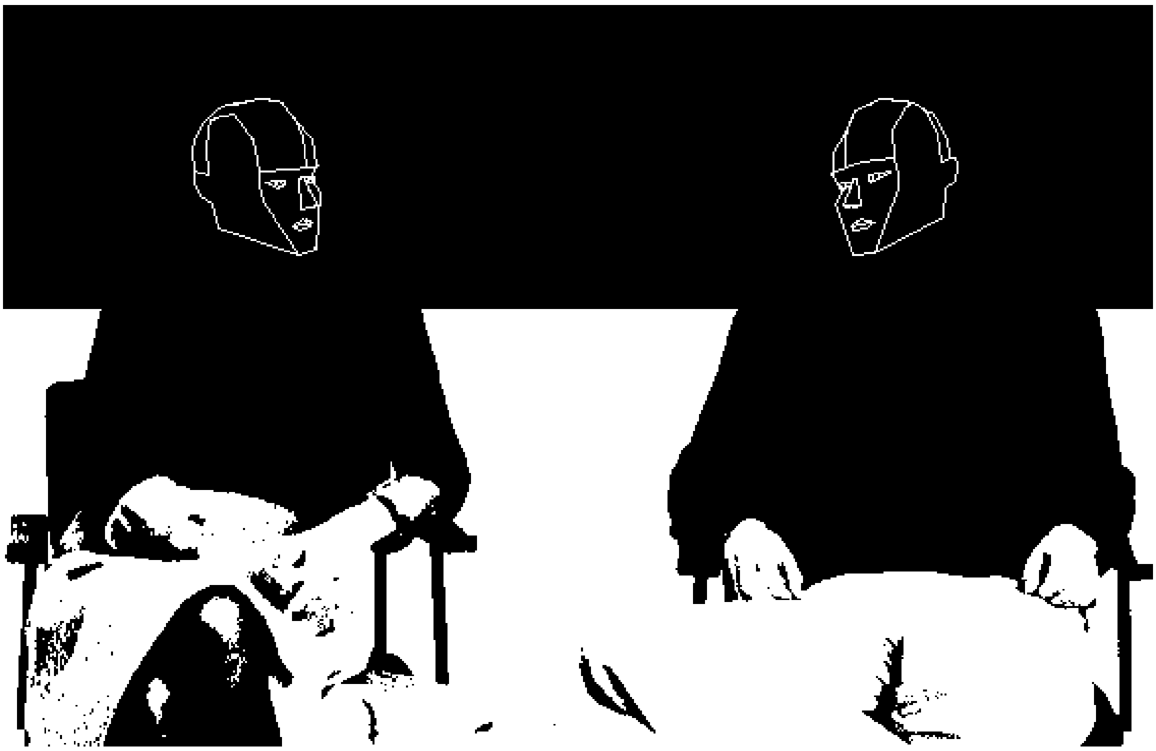

Figure 2: Screenshot from a head animation sequence of a mixed-sex dyadic interaction.

The relevant developments came in the early 90s with the availability of graphics workstations and sophisticated animation tools, such as the legendary Silicon Graphics (SGI) Indigo© and the animation program SoftImage3D©, which both played an essential role in the animations for “Jurassic Park.” Some external funding allowed me to purchase these nice toys and to develop a new generation of computer assisted movement transcription (Bente, Krämer, Petersen, & de Ruiter, 2001; Bente, Petersen, Krämer, & de Ruiter, 2001).

Although data quality again improved dramatically with regard to resolution as well as accuracy, the notation procedure was still very labor intense, requiring numerical data entries for each frame, each body part and the three spatial dimensions. Yet, being labor intense does not render a scientific work process useless or obsolete. In fact, only if a more economic procedure issues the same data quality with less effort should it replace the more labor intense one. I very much subscribe to the claim that for scientific methods, effectivity comes before efficiency. We cannot sacrifice information and data quality for measurement convenience. If the principle for optimization is evident, though, it is sometimes only a matter of time for technologies to offer more efficient solutions to the problem. This actually happened in the first decade of this century with affordable computer and graphics power and flexible software tools for avatar design and character animation. Together with my team I decided to switch to the software MotionBuilder ©, which is still our major modelling and animation frontend now at MSU, as it also easily binds to motion capture technologies as described below.

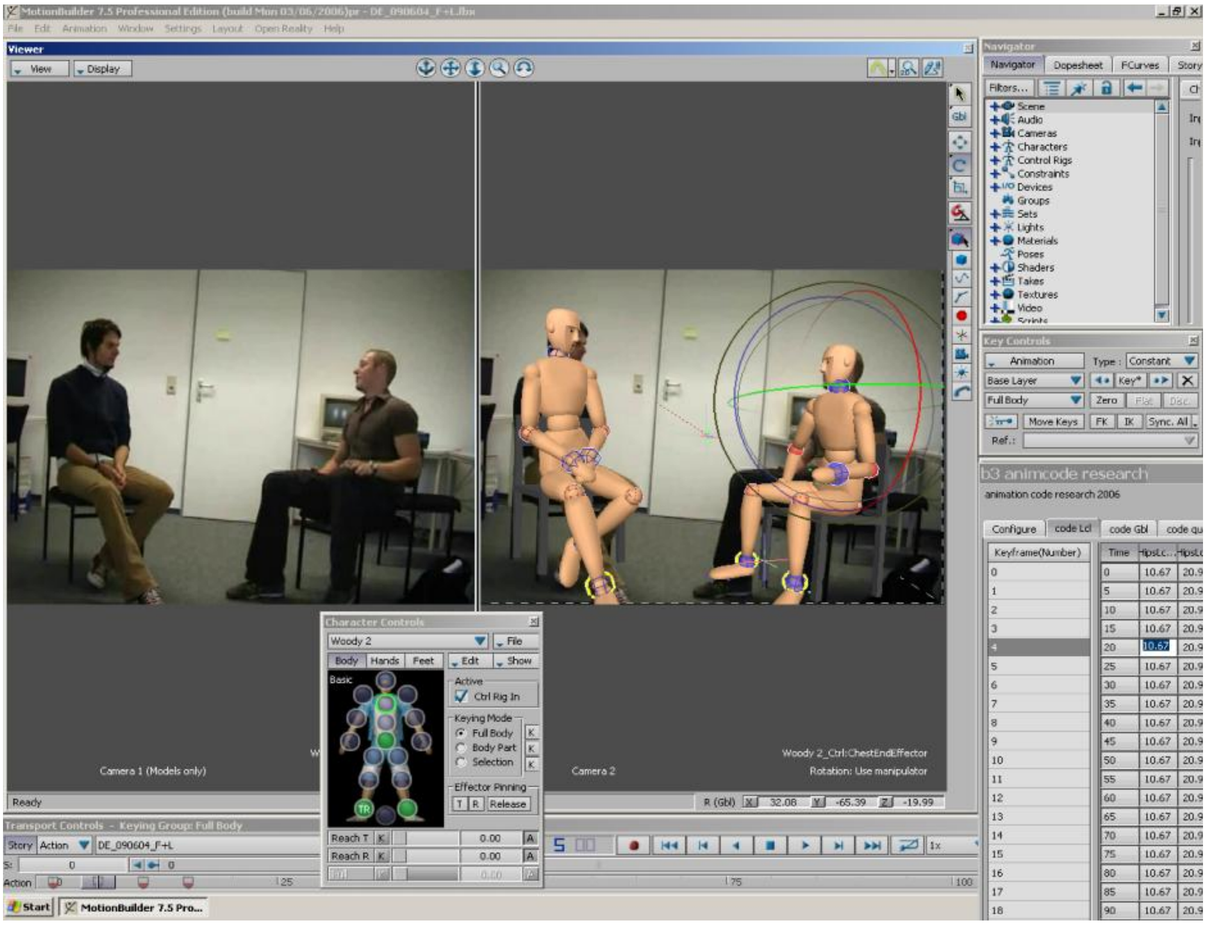

Figure 4: Screenshot of the MotionBuilder transcription interface.

The software has a special interface for Python plugins, which makes it easily adaptable to the researcher’s needs. Our first effort was to include a synchronized video player into the GUI, in a way that the avatar animations could be overlaid to the video frames. Figure 4 shows the GUI of the resulting transcription platform (cf. Bente, Leuschner, Al-Isa, & Blascovich, 2010).

The major advantage as compared to the Softimage© solution was that the movement of each body part could be controlled by means of a three dimensional sphere that could be rotated through mouse movements, picking either a single rotation dimension or combinations of any dimensions. To allow for unobscured visibility of the interactants, we integrated two video windows, one showing the recorded interaction without overlay. Data quality and transcription efficiency could be further improved this way. For the analysis of unstandardized video recordings of human movement in natural settings, where direct measurement via motion capture is not an option (e.g., when analyzing politicians' behavior in TV debates), the transcription rationale might still be without alternative.

Automatic Measurement with Full Body Motion Capture

Although computer-assisted transcription provided continuous data with high resolution and accuracy, it still depends on large time investments of human coders. For the sake of higher efficiency, automatic measurement procedures have been proposed as an alternative to transcription. For instance, Ramseyer and Tschacher (2011) introduced the so-called Motion Energy Analysis (MEA). MEA extracts general motor activity by essentially quantifying pixel changes between pre-filtered sequential video frames. In a newer approach, the authors also introduced a method to separate body and head movement within MEA (Ramseyer & Tschacher, 2014). Yet, resulting data protocols still only offer information about general movement activity and lack details about the postural dynamics of the various body parts. As mentioned above, machine learning algorithms extracting full body motion from webcam or other standard video recordings are rapidly developing. Keeping a close eye on this progress, I dare to say that they are not there yet. For this reason, I welcomed the advent of affordable motion capture technology in the last decade. Motion capture issues detailed protocols of body movement including rotation and translation information for all joints (cf. Bente, Roth, Dratsch & Kaspar, 2016; Poppe, van der Zee, Heylen, and Taylor, 2014).

Resulting data sets allow researchers to quantify postural dynamics of single body parts (e.g., head nods, head orientation, hand gestures, body lean, etc.) and also to aggregate motor activity across body parts (e.g., overall gestural activity). Different technologies are available. They either use passive markers (infrared reflecting buttons applied to a special full body suit) captured by special cameras (e.g., Optitrack: http://optitrack.com/), or active markers (using accelerometers and gyros also applied to special data suits) with no need of cameras at all (e.g., Rokoko: https://www.rokoko.com/en/). Newer systems include markerless devices (e.g. Organic Motion: https://tracklab.com.au/organic-motion/) using multi-camera high resolution video input. All these systems have advantages and downsides. Inertial tracking in contrast to the others is independent from the environment, as it needs no cameras. Yet, the position data shows considerable drift over time. Markerless systems are the least obtrusive for study participants, but are also less reliable with regard to hand and head rotations.

Right now in the CARISMA lab (Center for Avatar Research and Immersive Social Media Applications) at MSU, we use an Optitrack 24-camera (Prime 13) passive marker system, which simultaneously captures the movement of even small groups of people in each of two separate rooms. Calibration, capturing, and storage are performed by the software Motive©, which provides direct feedback using a standardized avatar. Movement data can be streamed in real time to the animation software Motion Builder and attached to a diversity of different avatars carrying out the actors' behavior on a neutral background or a specially designed virtual environment. Figure 5 shows a capture setup in our lab with a single person recording and the MotionBuilder representation in the foreground.

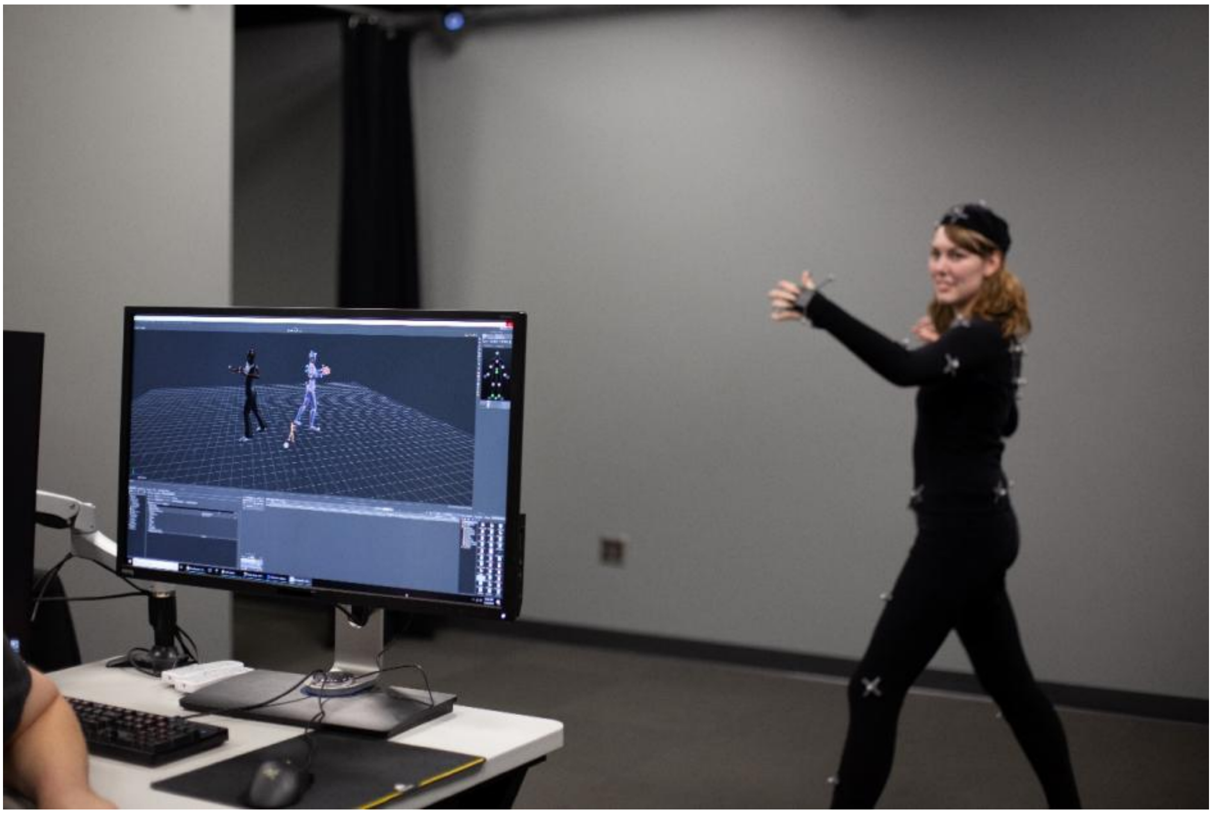

Figure 5: Single person motion capture setup in our CARISMA lab. The computer screen shows the Motion Builder GUI with the imported Motive character attached to an avatar.

Analyzing Descriptive Movement Protocols

It is important to stress that direct measurement via motion capture as well as computer-aided time-series notation result in the same type of data, i.e., protocols, which consist of time stamped entries with six Degrees of Freedom (DOF) for each of the 15 joints of a human body model. The entries comprise three translation values, i.e., the Cartesian coordinates for the 3D position in space, and three values describing their rotation around the three axes, typically in Euler angles. Time stamps allow us to synchronize the protocols of the interaction partners and to extract information about activity patterns, postural similarities or temporal contingencies within and across partners and movement dimensions. Based on early software developments, rationale of the analysis tools introduced by the Bernese research group had already been developed (see Hirsbrunner, Frey, & Crawford, 1987), but the tools were implemented on graphics computers expected to go off the market soon (Tektronix 4052, 32 kB RAM). With the advent of the first IBM PC and MS-DOS operating system in the early 1980s, I decided to develop my own programming skills to advance the analysis tools within the new framework.

Software development continuously occurred over the last two decades resulting in a program suite provided by one of my doctoral students, Haug Leuschner, at the University of Cologne (Leuschner, 2009). The program APEx (Automatic Parameter Extraction of Nonverbal Parameters) can be downloaded from the http://www.apex-download.eu. The website also contains a version of his dissertation (Leuschner, 2009), which explains all the algorithms and contains an elaborate user manual for interactive parameter definition and data export. The program uses ASCII files as input, which contain the movement dimension as columns and the time line as rows. The software is set up to work with any kind of interaction time series compatible with this format. It also comes with a Python plugin, which allows us to export data from animation software Motion Builder (.FBX files) into the APEx input format. An important component of the software is a conversion routine for the rotation values, usually Euler angles, used in the FBX-files. In contrast to the flexion levels used in the Bernese System, Euler angles can be equivocal and do not correspond well to the geometry underlying human perceptions of postures and movements. To provide rotation information, which is compatible with human perception, APEx converts Euler angles into flexion levels equivalent to the Bernese Code (for details see Leuschner, 2009). Overall, the program offers 150 predefined parameters including individual activity levels, postures, symmetry and openness of extremities as well as dyadic parameters such as interpersonal distance, mimicry, and head orientation towards the partner, among others. Further parameters can be defined using a special interface. All parameters can be output as time series for either individuals or dyads in the sample, or as an aggregate file (.CSV) containing the averages for all parameters and individuals, which is ready for statistical analysis.

The separation of description and evaluation in the measurement operation affords the re-integration in the stage of analysis. For this purpose, I recently developed an interactive data-mining tool (MotionLab), which builds on the APEx output. The program allows to combine the dyadic movement protocols with continuous impression ratings of observers and to synchronize these time lines with original videos and the hereon-based animation sequences on a frame-by-frame basis. It is an exploratory tool, which facilitates the search for critical events and recurrent patterns in the course of dyadic interactions. Figure 6 depicts a screenshot of the program. It shows a selected dataset from a study in progress in which we evaluated the rapport in male and female interactions across different cultures. The upper-left window contains the actual file selection. The upper middle holds the Windows Mediaplayer with the stimulus animation. A second player can open the original video if data were collected via time-series natation. The upper right window shows time clips selected by the researcher based on striking features in the time line either of the behavioral or the rating data. The lower-left window displays the timeline of the selected parameter and the rating averages. The lower right window shows an expandable set of aggregate parameters, which are calculated separately for the marked clips. In this example, the upper time graph shows the proxemics variation (interpersonal distance between the hips) between the two interaction partners in two different dyads. The lower time graph shows the according average ratings of rapport for these two dyads collected as a real time response measure. The data suggest that shorter distance goes with higher perceived rapport. This is indeed what we found in a first result of regression analysis conducted for various parameters, including suspected cues, such as partner-oriented head rotation and motor mimicry (Tickle-Degnen & Rosenthal, 1994).

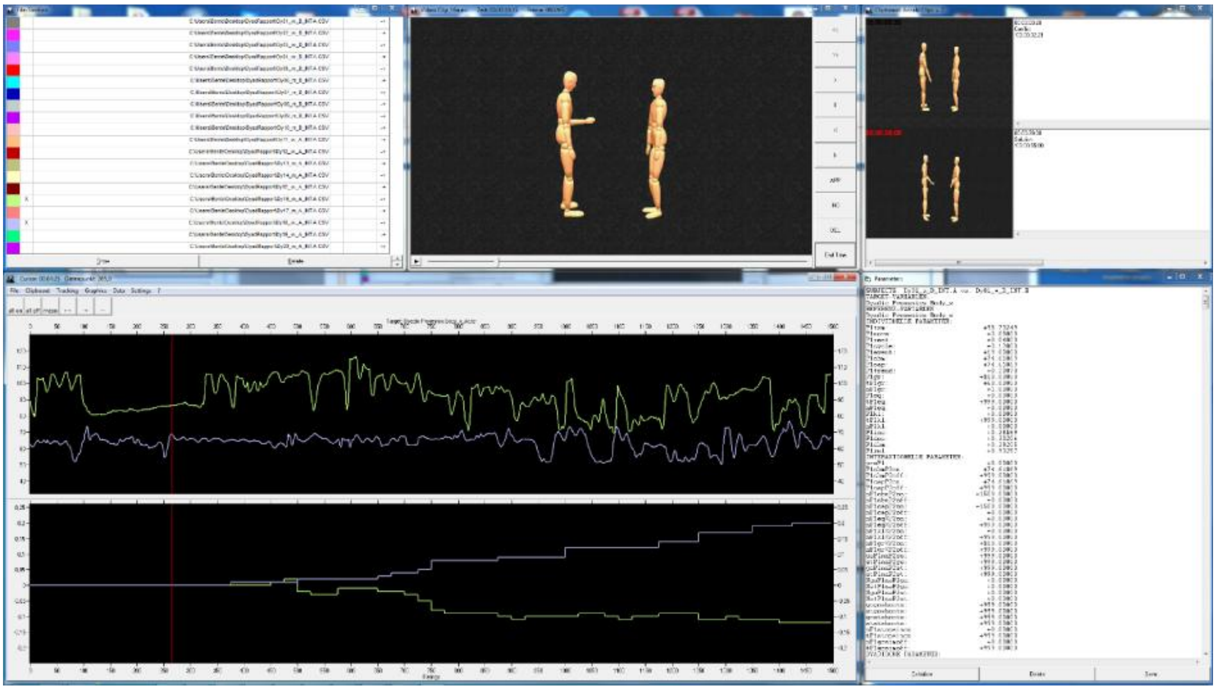

Figure 6: User Interface of the data-mining program MotionLab.

Addressing the Challenge of Experimental Control

Early Attempts to Solve Problems of Confound

Different methods have been proposed to solve the problem of confound in the stimulus generation for experimental person perception studies (cf. Bernieri, Davis, Rosenthal, & Knee, 1994). As mentioned above one of the most critical confounds is the one between aspects of physical appearance and behavior, in particular, if appearance features disclose a stereotype-relevant category membership of the target, such as gender, culture, age or race. This problem can hardly be overcome by randomization if group membership itself is a factor of interest. Berry, Kean, Misovich, and Baron (1991) discuss the use of so-called point light displays as a possibility to study the perceptual effects of human movement "…in the absence of potentially contaminating cues such as physical appearance" (p. 82). Point light displays reduce body movements to a simple image of white markers highlighting key joints against a black background. A disadvantage of point light displays, however, is that they only convey dynamic movement information but not static cues, such as postures and orientations (see Runeson & Frykholm, 1993). An alternative method has been introduced as "quantization" (see Berry, Kean, Misovich, & Baron, 1991). Quantization techniques degrade video images to rougher mosaic patterns, thus obscuring some structural features while preserving the dynamic information. In fact, quantization is not sufficient to completely obscure gender and culture, which might be relevant with regard to stereotype activation and related judgment biases (see for instance stimulus examples in Bernieri et al., 1994). More importantly though, the stimulus material is still video and not data which limits the possibilities for analysis and experimental control. Again emergent technologies proofed to be very helpful in solving the inherent methodological problems (see figure 1). Limitations of point light displays as well as quantization procedures can be overcome by using computer animations of virtual characters, so-called avatars. Based on physicalistic notation or motion capture data avatars can display the whole range of postures and movements while their appearance can be controlled with regard to gender, culture, and other impression-relevant appearance features. Furthermore, the nonverbal behavior of virtual characters can be systematically varied by editing the motion protocols.

First Avatar Studies: Method Validation

Fist studies using these character animations as stimuli in person perception experiments were encouraging (Bente, Feist, & Elder, 1996). It became evident that such data-driven motion simulations not only enabled a free combination of appearance and behavior, as in the example above, but also opened up enormous degrees of freedom in the variation of individual behavioral aspects through targeted intervention in the numerical behavioral protocols. In fact, our early approach had considerable limitations in terms of representation realism and the bandwidth of the detected or generated interaction behavior. These animations were constrained to head movements and to a very simple, uniform wireframe, which could be overlaid to black and white images of static bodies (see Figure 2 for an example). With constant advances in 3D modeling and animation, virtual actors became more and more realistic in terms of both appearance and movement. Despite these technological advancements the central questions remained to be answered, to what extent computer simulations lead to realistic socio-emotional effects and produce impressions that are comparable to observations of real life nonverbal behavior or full body video recordings. To answer this question I conducted a series of validation studies, in which the impressions of video recordings of dyadic interactions were compared with those of 3D computer animation. At that time, the simulations were still based on transcriptions of the video documented movement behavior using the Bernese System. The behavioral logs, however, were animated using a professional animation software (SoftImage3D). The experimental stimuli comprised the original video as well as a surface model and a wireframe model of the avatar reconstruction (see figure 7). We measured observer impressions of both partners with a 20-item adjective list as well as visual attention with eye tracking. Results were striking. As expected, eye tracking showed marked differences between video and the animations, which did not show facial variation. Yet, the impression profiles for a variety of judgments (e.g., active, distressed, interested), also shown in Figure 7, were nearly identical across all three conditions (cf. Bente, Krämer, Petersen, & de Ruiter, 2001). We replicated this result with another six dyads (Bente, Petersen, Krämer, & de Ruiter, 2001). From this point on, I was confident that we had a new method for computational experimentation, which represented a perfect compromise between ecological validity of the stimuli as well as unprecedented degrees of experimental control.

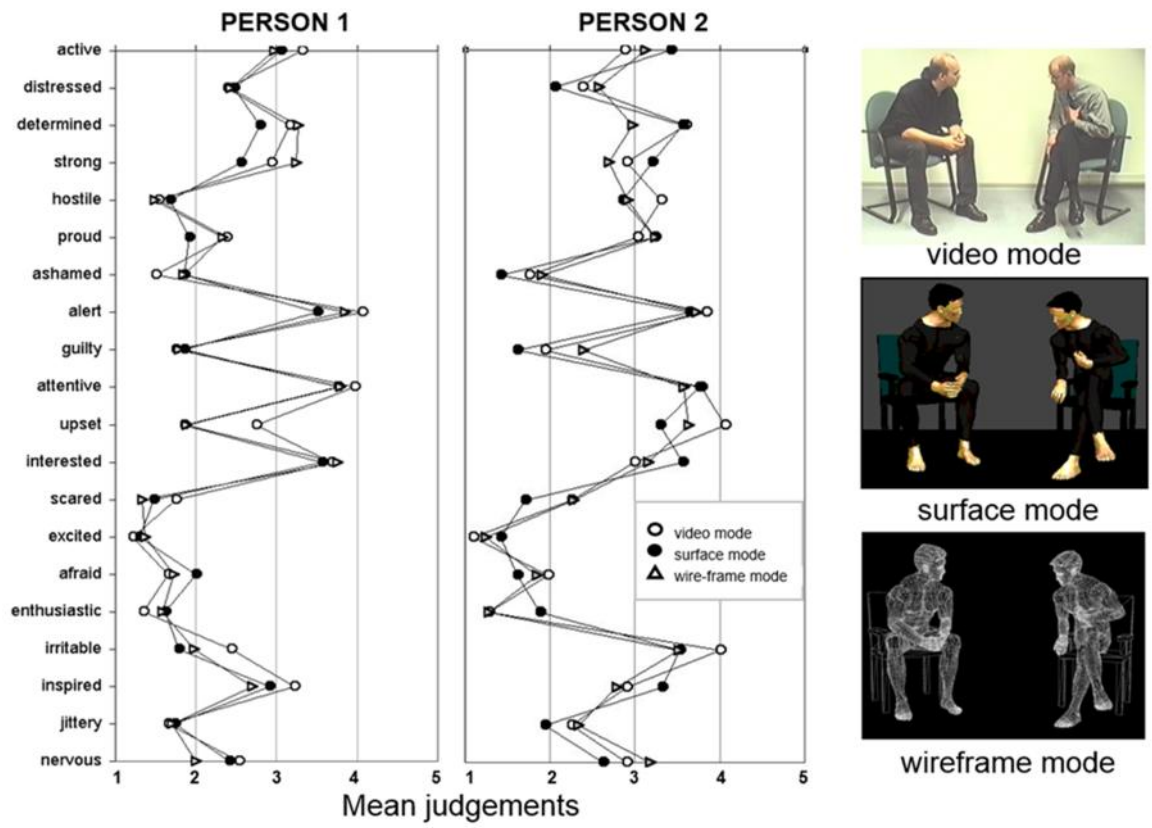

Figure 7: Impression ratings under three stimulus conditions from the first method validation studies (cf. Bente et al. 2001, a,b).

Since the early 2000s, we have used this methodology for numerous studies, lately replacing the time consuming transcription process by the motion capture procedures described above. In fact, this increases resolution as well as accuracy and reduces workload. Yet, it changes neither the principle of behavior representation nor the computer animation rationale. In the following section, I will present a few study examples that demonstrate the unique potential of the avatar technology in different study types. These include more *observation studies* as well as *real time interaction studies* with humans represented through avatars.

Stereotype Influence in Observation Studies

The influence of stereotypes on judgments of behavior has been discussed as a crucial problem in evaluation studies. Solutions depend on the research question. If stereotype influence is to be eliminated, the solution is to use neutral avatars, which conceal any information about relevant category memberships of the targets. If the stereotype influence were part of the research question the solution would be to vary cues to group membership systematically.

In a study of the first type, we investigated universals and cultural specificities in impression formation from nonverbal interaction behavior across three cultures: Germany, the United States and the United Arab Emirates (Bente, Leuschner, Al-Issa, & Blascovich, 2008). We used a neutral avatar resembling a wooden artist manikin to eliminate hints to culture and ethnicity. Animations again were based on computer-assisted notation with MotionBuilder as described above (see Figure 3). Figure 8 shows the examples of the animation derived from the original video recordings.

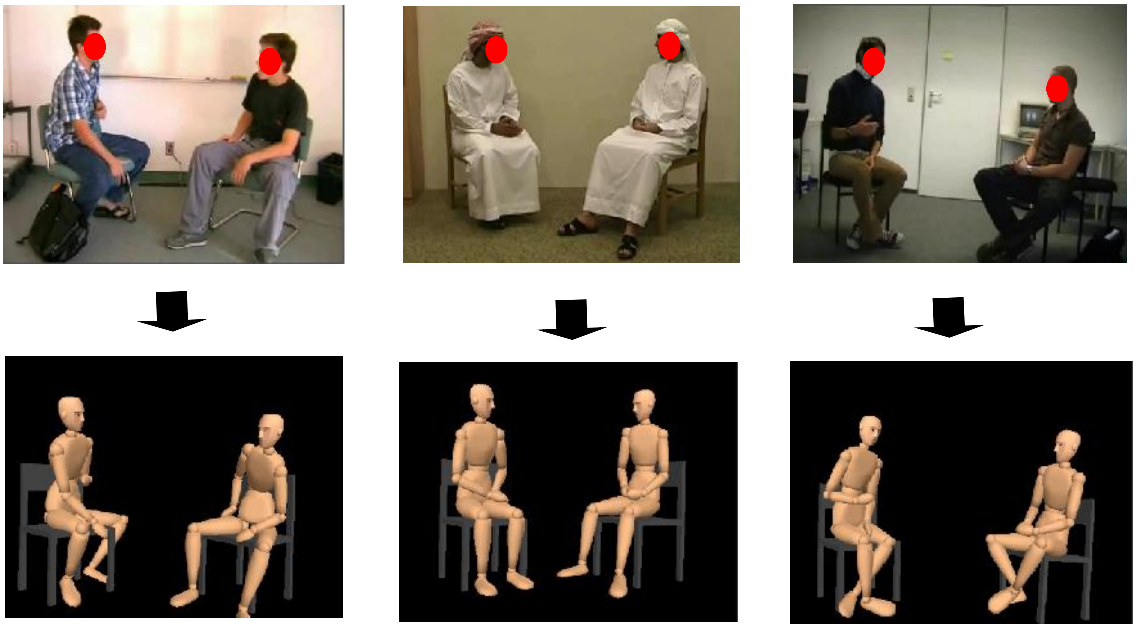

Figure 8: Examples of animation sequences in the cross-cultural study using a neutral manikin-like avatar (from: Bente, et al., 2008).

The treatment check revealed that observers were not able to identify the culture of origin of the interactions. Impressively, dominance cues seem to be universal. Dominance ratings were highly correlated across observers from all countries, while valence ratings showed marked differences. Microanalysis of the nonverbal data protocols also revealed that the dominance judgments in the three cultures were driven by the same nonverbal cues including raised head positions, partner directed head orientation, and opening of the upper extremities.

In studies of the second type, we investigated the cultural judgment biases in emotion inferences. (Bente, Dratsch, Kaspar, Häßler, Bungard, & Al-Issa, 2014). We used actor performances expressing anger and happiness recorded with a 12-camera Optitrack motion capture system (see Figure 4) to animate typical exemplars of the Western and Arab culture (see Figure 9). In contrast to our hypothesis, there was no interaction effect between culture of observers and culture of avatars. However, there was a general negativity effect for the European avatar. An additional treatment check revealed that the effects of the physical appearance, which was negative for the European avatar, had overridden the effects of the emotional expression. The unappealing European avatar was just more likely to be associated with negative (anger) than positive (happiness) emotions. This result points to the relevance of the static features of the virtual characters, a result that has far-reaching implications for the use of avatars in shared virtual realties (cf. Fong & Mar, 2015).

In a further study, based on a minimal group paradigm (Bente, et al., 2016), we used neutral manikin avatars characterized as ingroup or outgroup by means of jersey colors, either matching the ones of the participants or not. Figure 10 shows the examples of the stimuli. Surprisingly we found significant differences in the responses to anger versus happiness stimuli within the ingroup condition but not within the outgroup condition. The results suggested a more critical evaluation of the ingroup stimuli, which we interpreted as an attentional effect caused by the greater salience of these stimuli (cf. Bente et al., 2016).

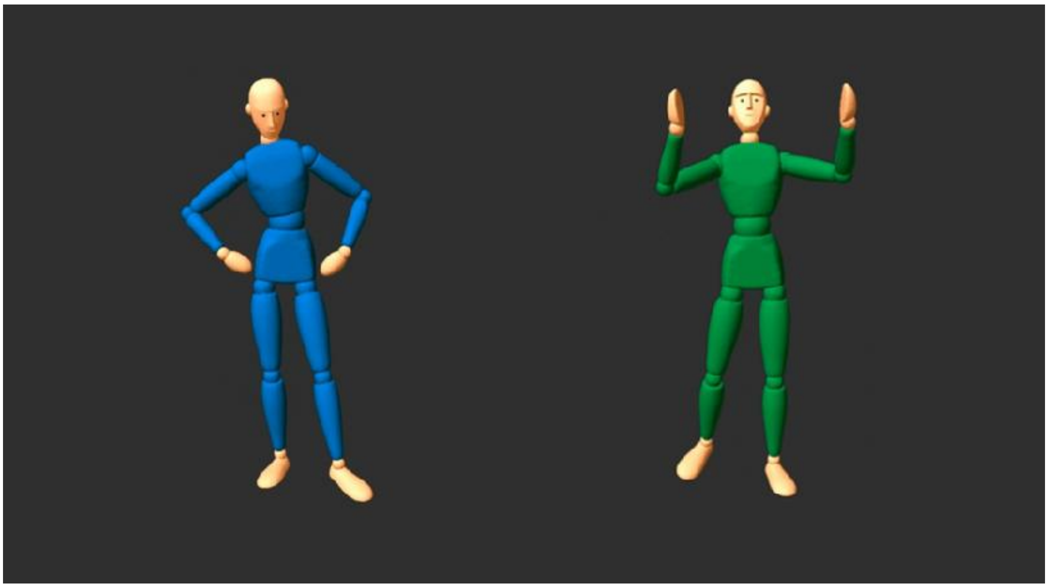

Figure 9: Culture-Typical Avatars used in a Cross-Cultural Emotion Recognition Study (cf. (Bente et al., 2014).

Figure 10: Avatars used in a minimal group paradigm study on emotion recognition (Bente, et al., 2016)

Interactive Paradigms and Transformed Behavior

One of the major advantages of motion capture as compared to physicalistic notation is the possibility to use avatars and manipulate their physical appearance as well as their nonverbal behavior online in ongoing interactions (cf. Bailenson & Beall, 2005). We presented a first study on online manipulations of behavior over a decade ago (Bente, Eschenburg & Krämer, 2007). In this study, we used a desktop-VR system displaying the upper body of two interactions as simple avatars (stick figures with pronounced eyes) on screen. Movement was captured by a Polhemus tracker, and hand motion by means of Cybergloves. In addition, gaze behavior was measured with our own eye-tracking device. While all nonverbal channels were transmitted in real time, eye gaze of one partner was replaced by artificial gaze behavior controlled by a computer algorithm. Figure 11 shows an example of the study setup. Systematic variation concerned the duration of directed gaze (looking into the partner's face and enabling eye contact). Results showed that increased visual attention had a significant impact on partner liking. In a follow up brain imaging study, using avatars controlled by a similar algorithm, we also showed that the processing of gaze direction and gaze duration recruits distinct neural systems (Kuzmanovic, et al. 2009), with gaze duration being associated with differential neural activity in the medial prefrontal cortex (mPFC), which is an area relevant to the construction of social meaning. In fact, the use of highly controllable avatars proved to be essential for the study of nonverbal communication in the context of social neuroscience with which I enjoyed an intense collaboration over the last two decades (see Vogeley & Bente, 2010).

Figure 11: Desktop avatar-communication platform used in previous studies on synthetic gaze behavior (Screenshot shows interlocutors during calibration phase displaying their own avatars)

Interactive avatar studies making use of manipulations of physical clues to identity and group membership are a focus of current research projects investigating the socio-emotional effects of motor synchrony, which is supposed to build the basis for collaboration and to contribute to group entitativity and pro-social attitudes. In a series of studies, we are currently investigating the influence of motor synchrony on intergroup trust (see Tamborini, et al. 2018). In a study in progress, for instance, we try to eliminate racial biases and foster outgroup trust using our real-time avatar interaction platform. Participants perform a Tai-Chi exercise together where they are instructed to synchronize their movement as well as possible. The partners are displayed as avatars either representing the ingroup (in our sample's case, Caucasian) or the outgroup (African-American) relative to the actual race of the participant. Figure 12 shows two images of the study setup. We expect interaction with an outgroup member to increase outgroup trust and that this effect is moderated by the level of motor synchrony achieved over time. Our next steps will include systematic variation of synchrony as seamlessly inserted by a computer algorithm.

This approach not only exemplifies the unprecedented experimental possibilities, it also illustrates relevant applications of the technology in training and educational scenarios. It is important to note, however, that plasticity of appearance and behavior within shared virtual environments of the future also bares risks of fraught and manipulation. The analysis of these risks and of potential social validation strategies is a further task for a future-oriented communication science.

Figure 12: Setup of an ongoing study on the prosocial effects of motor synchrony. Left: Joint TaiChi performance of a white participant with a neutral avatar. Right: Start T-pose of the white participant and a black avatar.

Conclusions and Future Prospects

In this chapter, I tried to make a statement for nonverbal communication research as a hard science. This implies, in my view, a considerable investment into objective measurement, which allows for a comprehensive and detailed representation of the physical properties of human behavior (cf. Barker, 1964). Independent from this process we then can collect data representing the social effects of nonverbal behavior, including verbal evaluations, behavioral responses or even neural response patterns. Meaningful cues, i.e., the socially significant variations in the complex stream of human movement behavior, are identified as events in the timeline, which cause intersubjective, common responses in an audience, either verbal, behavioral or neural. This view of communication effects is increasingly represented in communication neuroscience aiming at the identification of effective messages, e.g., in health communication (Imhof, Schmälzle, Renner, & Schupp, 2017) or political speeches (Schmaelzle, Hacker, Honey, & Hasson, 2015), and I am convinced that the so-called inter-subject-correlation (ISC) paradigms are useful in nonverbal communication research as well. I see a particular challenge in adding neural process data to the time-series protocols of nonverbal behavior and continuous observer ratings to search for objective correlates of the subjective impressions.

Technological progress in the area of Virtual Reality (VR), including affordable, markerless motion capture technology, avatar modelling tools, real-time character animation software, and immersive display technologies, will dramatically improve the methodological basis of our research. Taking advantage of these developments, we are currently working on a so-called hybrid-avatar-agent technology (Roth, Latoschik, Vogeley, & Bente, 2015), which allows control of the physical appearance of the avatars as well as the rendered behavior in real-time by means of computer algorithms. Moreover, advancements in machine learning can be expected in the near future that will enable the automatic extraction of human movement from standard video recordings (see Sarafianos, Boteanu, Ionescu, & Kakadiaris, 2016). All these developments will give us access to large and detailed data sets of human communication behavior and unprecedented experimental control over nonverbal cues as represented by avatars or virtual agents within shared virtual environments.

New tools are required, though, to perform the data mining in the huge and highly complex interaction data bases we expect. These methods are not readily available from the shelf. Therefore, I strongly recommend that next-generation researchers in the field acquire some programming knowledge to be able to adapt their tools to their needs. As I tried to show, technology development taking place outside the realm of communication science can enrich our methodological arsenal, promote new research paradigms, and advance our knowledge. Yet, the use of it requires some technical expertise and particularly some programming skills. Although I am well aware of the increasing publication pressure and the urge for external funding, I would like to encourage next-generation researchers to develop those skills and gain some autonomy with regard to tailoring their measurement instrumentation and analysis tools. In fact, I am delighted to see this development in the field already and I am impressed by the sharing attitude of the scientists within open source communities such as GitHub.

References

Argyle, M. (1975). *Bodily communication.* London: Methuen.

Aronson, E. (1972). *The Social Animal*. San Francisco: Freeman.

Asch, S.E. (1948. The doctrine of suggestion, prestige and imitation in social psychology. *Psychological Review,* 55, 250-276.

Bailenson, J. and Beall, A. (2005). Transformed social interaction: Exploring the digital plasticity of avatars. In: R. Schroeder and A.-S. Axelsson (Eds.). *Avatars at work and play: Collaboration and interaction in shared virtual environments*. London: Springer.

Bente, G. (1989). Facilities for the graphical computer simulation of head and body movements. *Behavior Research Methods, Instruments, & Computers, 21*, 455-462.

Bente, G., Donaghy, W.C., & Suwelack, D. (1998). Sex differences in body movement and visual attention: An integrated analysis of movement and gaze in mixed-sex dyads. *Journal of Nonverbal Behavior, 22,* 31-58.

Bente G., Dratsch T., Rieger D., Al-Issa A. (2014). Emotional contagion with artificial others - Effects of culture, physical appearance and nonverbal behavior on the perception of positive/negative affect in avatars. In Meiselwitz G. (Ed.) *Social Computing and Social Media. SCSM 2014. Lecture Notes in Computer Science,* vol. 8531 (pp. 411–420). Berlin: Springer.

Bente G., Eschenburg F., Krämer N.C. (2007). Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations. In Shumaker R. (Eds.) *Virtual reality. ICVR 2007. Lecture notes in computer science*, vol. 4563 (pp. 185-194). Berlin: Springer.

Bente, G., Feist, A., & Elder, S. (1996). Person perception effects of computer-simulated male and female head movement. *Journal of Nonverbal Behavior, 20*, 213-228.

Bente, G., Frey, S., & Hirsbrunner H.P. (1984) Analyse nonverbaler Interaktion als Mittel der Prozessforschung. In U. Baumann (Ed.) *Psychotherapie Makro-/Mikroperspektiven* (p. 240-264). Göttingen: Hogrefe.

Bente, G., Kraemer, N. C., Petersen, A., & de Ruiter, J. P. (2001a). Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior, 25*, 151–166.

Bente, G., Leuschner, H., Al-Issa, A., & Blascovich, J. J. (2010). The others: Universals and cultural specificities in the perception of status and dominance from nonverbal behavior. *Consciousness and Cognition*, 19, 762–777.

Bente, G., Petersen, A., Kraemer, N. C., & De Ruiter, J. P. (2001b). Transcript-based computer animation of movement: Evaluating a new tool for nonverbal behavior research. *Behavior Research Methods, Instruments, & Computers*, 33(3), 303–310.

Bente, G., Roth, D., Dratsch, T., & Kaspar, K. (2016). Emotions of my kin: disambiguating expressive body movement in minimal groups. *Culture and Brain*, 4, 51-71.

Bente, G., Rüggenberg, S., Krämer, N.C., & Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Human Communication Research, 34*, 287-318.

Bente, G., Senokozlieva, M., Pennig, S., Al-Issa, A., & Fischer, O. (2008). Deciphering the secret code. A new methodology for the cross-cultural analysis of nonverbal behavior. *Behavior Research Methods*, 40, 269–277.

Bernieri, F. J., Davis, J. M., Rosenthal, R., & Knee, R. C. (1994). Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin*, *20*, 303–311.

Berry, D. S. (1990). The perceiver as naive scientist or the scientist as naive perceiver? An ecological view of social knowledge acquisition. *Contemporary Social Psychology, 14*, 145–153.

Berry, D. S., Kean, K. J., Miscovich, S. J., & Baron, R. M. (1991). Quantized displays of human movement: A methodological alternative to the point-light display. *Journal of Nonverbal Behavior*, *15*, 81–97.

Cappella, J. N. & Palmer, M. T. (1990). The structure and organization of verbal and non-verbal behavior: Data for models of production. In H. Giles, W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 141-161). Chichester: John Wiley & Sons.

Donaghy, W. C. (1989). Nonverbal communication measurement. In P. Emmert & L. Barker (Eds.), *Measurement of Communication Behavior* (pp. 296-332). White Plains, N. Y.: Longman.

Fong, K.I. & Mar, R.A. (2015).What does my avatar say about me? Inferring personality from avatars. *Personality and Social Psychology Bulletin, 41*, 237-249.

Frey, S., Hirsbrunner, H. P., Florin, A., Daw, W. & Crawford, R. (1983). A unified approach to the investigation of nonverbal and verbal behavior in communication research. In W. Doise & S. Moscovici (Eds.), *Current issues in European social psychology* (pp. 143-199). Cambridge: Cambridge University Press.

Frey, S., Hirsbrunner, H.-P. & Bieri-Florin, A. (1979). Vom Bildschirm zum Datenprotokoll: Das Problem der Rohdatengewinnung bei der Untersuchung nichtverbaler Interaktion. *Zeitschrift für Semiotik, 1,* 193-209.

Galliker, M. (1990). Ist die psychologie eine exakte wissenschaft? Zur befolgung der äquivalenzvorschrift beim messen am beispiel des „Berner System“. *Psychologische Beiträge, 32* , 225-248.

Hirsbrunner, H. P., Frey, S. & Crawford, R. (1987). Movement in human interaction. Parameter formation and analysis. In A. W. Siegmann & S. Feldstein (Eds.), *Nonverbal behavior and communication* (pp. 99-140). Hillsdale, N.J.: LEA.

Hut, P., & Sussman, G. J. (l987). Advanced computing for science. *Scientific American,* 257, 144-153.

Imhof, M.A., Schmaelzle, R.H., Renner, B. Schupp, H.T. (2017). How real-life health messages engage our brains: Shared processing of effective anti-alcohol video. *Social Cognitive and Affective Neuroscience 4,* 1188–1196.

Kuzmanovic, B., Georgescu, A.L., Eickhoff, S.B., Shah, N.J., Bente, G., Fink, G.R., & Vogeley, K. (2009). Duration matters: Dissociating neural correlates of detection and evaluation of social gaze. *Neuroimage, 46*, 1154-1163.

Leuschner, H. (2013). Microanalysis of nonverbal communication: Development of a nonverbal research method using high-performance 3D character animation. Doctoral dissertation, University of Cologne, Germany. Retrieved from http://www.apex-download.eu

Lewis, R. J., Derlega, V. J., Shankar, A., Cochard, E. & Finkel, L. (1997). Nonverbal correlates confederates´ touch: Confounds in touch research. *Journal of Social Behavior and Personality, 12* , 821-830.

Patterson, M. L. (1990). Functions of nonverbal behavior in social interaction. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology* (pp. 101-120). Chichester: John Wiley & Sons.

Poppe, R., Van Der Zee, S., Heylen, D., & Taylor, P. J. (2014). AMAB: Automated measurement and analysis of body motion. *Behavior Research Methods, 46*, 625-633.

Puccinelli, N. M., Tickle-Degnen, L., & Rosenthal, R. (2003). Effect of dyadic context on judgments of rapport: Dyad task and partner presence. *Journal of Nonverbal Behavior, 27*, 211-236.

Ramseyer, F. & Tschacher, W. (2011).Nonverbal synchrony in psychotherapy: Coordinated body-movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology, 79*, 284–295.

Ramseyer, F. & Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Frontiers in Psychology, 5* (979).

Roth, D., Latoschik, M.E., Vogeley, K., & Bente, G. (2015). Hybrid avatar-agent technology – A conceptual step towards mediated "social" virtual reality and its respective challenges. i-com, 14(2), 107-114. doi: 10.1515/icom-2015-0030

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*, 585–615.Sarafianos, N., Boteanu, B., Ionescu, B. & Kakadiaris, I.A. (2016). 3D Human pose estimation: A review of the literature and analysis of covariates. Computer Vision and Image Understanding, 152, 1-20.

Schmaelzle, R.H. Hacker, F.E.K., Honey, C.J., Hasson, U. (2015). Engaged listeners: Shared neural processing of powerful political speeches. *Social Cognitive and Affective Neuroscience, 10*, 1137–1143.

Tamborini, R., Novotny, E., Prabhu, S., Hofer, M., Bente, G., Grall, C., Klebig, B., Hahn, L., Slaker, J., Ratan, R. (2018). The effect of behavioral synchrony with black or white virtual agents on outgroup trust. *Computers in Human Behavior, 83*, 176–183.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry, 1*, 285–293.

Vogeley, K., & Bente, G. (2010). "Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Networks, 23*, 1077-1090.