The Challenge of Constructing Psychologically Believable Agents

Schönbrodt, F. D., & Asendorpf, J. B. (2011). The challenge of constructing psychologically believable agents. *(c) Journal of Media Psychology: Theories, Methods, and Applications*, *23*, 100–107. doi:10.1027/1864-1105/a000040

Author Note

Felix D. Schönbrodt, Department of Psychology, Humboldt University Berlin, Germany; Jens B. Asendorpf, Department of Psychology, Humboldt University Berlin, Germany.

Felix D. Schönbrodt is now at Department of Psychology, Ludwig-Maximilians-University Munich, Germany.

Correspondence concerning this article should be addressed to Felix Schönbrodt, Leopoldstr. 13, 80802 München, Germany. Email: felix.schoenbrodt@psy.lmu.de. Phone: +49 89 2180 5217. Fax: +49 89 2180 3000.

Summary

Embodied conversational agents (ECAs) are designed to provide a natural and intuitive communication with a human user. One current major topic in agent design consequently is to enhance their believability, often by means of incorporating internal models of emotions or motivations. As psychological theories often lack the necessary details for a direct implementation, many agent modelers currently rely on models that are rather marginal in current psychological research, or models that are created ad hoc with little theoretical and empirical foundations. The goal of this article is both to raise psychologists' awareness about central challenges in the process of creating psychologically believable agents, and to recommend existing psychological frameworks to the virtual agents community that seem particularly useful for an implementation in ECAs. Special attention is paid to a computationally detailed model of basic social motives that seems particularly useful for an implementation: the Zurich model of social motivation.

Keywords: embodied conversational agents, believability,

Zurich model of social motivation

For decades, humans had to speak the language of the machines when they wanted to communicate with them. Starting from the first mechanical calculating machines from the 17th century, where the operators had to understand the internal mechanisms to get valid results, this habit continued till the 1970s, when human still had to "talk" to computers via punch cards (Redin, 2007). A major shift in human-machine interaction took place in the 1980s with the introduction of graphical user interfaces and accompanying features like drag-and-drop, a trash can for the deletion of files, or a desktop (like in reality, cluttered with papers and files). None of these interface features is a necessity for the machine, but enhances usability by referring to well-known work flows from the physical world (e.g. "grabbing a file and throwing it into the trash can"). A preliminary culmination in the evolution towards a more user-centered communication has been reached in the efforts to develop human-like synthetic characters (embodied conversational agents, ECAs), whose purpose is to allow a natural verbal and non-verbal communication (Cassell, Churchill, Prevost, & Sullivan, 2000).

To allow humans a communication to computers that is as natural as possible, it is argued that ECAs should be built upon existing psychological theories about human communication, emotion, and motivation (Gratch & Marsella, 2004). But what can psychology contribute to build more believable embodied conversational agents? As Krämer, Bente, Eschenburg, and Troitzsch (2009) state, up to now there is only little exchange between the virtual agents community employing psychological theories and current psychological research. Although there have been several attempts to implement psychological theories into autonomous agents, many psychological theories lack the necessary details needed for a direct implementation. Software architects therefore have to do a lot of interpretation, extrapolation and "filling the gaps". Psychological theories largely differ in their level of detail, and during the process of implementation one soon will find some theories to be more suitable than others. Consequently, virtual agent architects

sometimes seem to choose their models mainly based on its computability. For example, the most frequently implemented model of emotion is the OCC model by Ortony, Clore, and Collins (1988), a theory which is not the most prominent one in psychological textbooks or current psychological research (Krämer, 2008). Furthermore, often specific psychological knowledge about communication and interactional processes is scarce, and it was concluded that one has to rely on the intuition of the designers and animators of ECAs (Cassell et al., 2000), or on ad hoc models which are based on dubious data or no data at all (cf. Krämer, 2008).

The current article addresses both the psychologists and the virtual agents community. On the one hand its goal is to increase psychologists' awareness about the shortcomings of many psychological theories when they are to be implemented in autonomous agents and to introduce some key challenges in the construction of ECAs. The example of the action selection problem is introduced to highlight an area where psychological input would be needed, but unfortunately is largely missing. On the other hand, psychological theories in the domains of interpersonal perception and behavioral synchrony are presented as a possible psychological input into the design of ECAs. Finally, the Zurich model of social motivation is presented as a theory that is both psychologically sound and provides detailed information about computational details, which eases potential implementations by the virtual agents community.

Believability

As a goal in the construction of ECAs is a flawless and smooth conversation with human interaction partners, enhancing the believability of ECAs generally is a sensible goal. But what is the scientific concept of believability, beyond the common sense of the word? No final definition of believability has been come up so far, and several fine-grained distinctions of different types of believability have been proposed (e.g., Rose, Scheutz, & Schermerhorn,

2010). Believability can be conceptualized both as a property of the ECA (e.g., special conversational abilities that raise the feeling of believability in most human interaction partners), as well as a property of the perceiver (e.g., the “hardened robotics researcher”, for whom it is very unlikely to ascribe mental states to any robot due to his intricate knowledge of the underlying mechanisms; Rose et al., 2010).

For the purpose of this article, we focus on properties of ECAs that contribute to the feeling of believability in a human perceiver, and we put forward following operational working definition of believability: Believability of an ECA is the extent to which human interaction partners can intuitively communicate with it by applying natural processes of human communication. The numerous processes involved can include the perception of a consistent personality of the ECA (Ortony, 2003), which is expressed through behavior that is consistent with the agents' goals or states of mind, as well as a consistency between verbal and non-verbal communication (De Rosis, Pelachaud, Poggi, Carofiglio, & Carolis, 2003). Internal models of emotion and motivation should be a promising way to achieve this consistency (see below). Furthermore, the consideration of natural processes of interpersonal perception should increase the feeling of believability. These processes, which all contribute to the impression of believability, are diverse aspects which might well be separately analyzed. However, we would argue that at a higher level of abstraction all these aspects contribute to a “general factor” of believability.

Thus, the general idea of believability is to enhance and ease human-machine interaction by using natural and intuitive codes of human communication. Humans have certain expectations about communicative signals, patterns, and reactions of interaction partners. Believability of an ECA means that humans can apply their usual mental models of communication.

Why should designers of ECAs want to increase the believability of their agents? The ultimate intentions of making agents believable can be very different and include for example (a) better user experiences in games, virtual drama, or arts ("suspension of disbelief", Bates, 1994); (b) higher impact of cybertherapy or training exercises (e.g. Beutler & Harwood, 2004); (c) higher external validity in virtual social psychology experiments (e.g. Blascovich, Loomis, Beall, Swinth, Hoyt, & Bailenson, 2002); (d) a more effective and robust transmission of information through the use of multiple communicative channels (e.g. Cassell, 2001); or (e) the simulation of psychological theories, when the evaluation of the agents' believability is the criterium of interest (Wehrle, 1998).

How can believability be achieved? We already mentioned several psychological processes above. Maybe the most discussed approach to believability is the implementation of emotions and motivations into agents (e.g. Becker-Asano & Wachsmuth, 2008; De Rosis et al., 2003; Gratch, 2008; Hudlicka, 2003; Ortony, 2003). However, it has been debated whether human-like *internal processes* are necessary at all, or whether the replication of *surface displays* is sufficient for a believable and effective communication. Krämer, Iurgel, and Bente (2005) argue that to reach the ultimate goal of an ECA - to make the human-computer interaction more intuitive and to manipulate the emotions of the user - it is not necessary to simulate internal processes in terms of an emotion model or motivation model. Instead they propose a conversational function model where emotional expressions are seen as purely instrumental and disconnected from the feeling of emotions (cf. Fridlund, 1991). In this alternative approach, non-verbal behaviors (like emotional displays) are directly chosen based on their known (and intended) effect on the user, without the need of a simulated emotional state of the agent. By doing so, the realized system would be more effective and easier to implement. In some restricted scenarios this alternative approach might be straightforwardly implemented. In more open scenarios (e.g., in a soft-skills business training simulation), these

agents, however, soon will reach their limits as situations will arise that are not covered by the internal database (Ortony, 2003). In such an application scenario, agents who are guided by more general underlying principles presumably produce more consistent and coherent behavior. Accordingly, Gratch and Marsella (2004) argue that internal processes and emotional expressions are closely linked, and that psychologically informed theories about internal processes should form the basis for communicative processes in interactive settings (see also Gratch, 2008).

To be clear, we do not argue that an implementation of human psychological mechanisms is mandatory to achieve an agent's believability – a newly developed "non-human" architecture of an ECA could as well accomplish the same goal, along Krämer et al.'s argumentation. However, models which simulate actual human psychological processes might be a good starting point for the task.

In the remainder some key challenges in the construction of ECAs are presented, and psychological theories are presented, whose implementation is promising for the increase of believability.

Action Selection and the Persistence of Behavior

One challenge every designer of agent architectures has to face is the development of a mechanism that decides which specific behavior the agent should initiate in the presence of multiple external or internal driving forces. This decision is also called the problem of "action selection", and two oversimplified solutions should illustrate the difficulty:

(a) pure dominance of motives: motives are in a fixed hierarchy, for example "flight" always dominates "eating", which in turn always dominates "mating". The lower priority motive only can be expressed if all higher priority motives are satisfied.

(b) "the winner takes it all": the motivation with the momentary highest activation (due to internal or external factors) gains control over behavior.

However, both approaches inadequately deal with the problem of persistence of behavior. For the following examples imagine the simulation of physiological needs in a simulated environment - namely hunger and thirst - which both rise continuously over time and have to be regularly satisfied. If thirst always dominates hunger, an agent with a slightly higher thirst motivation will stay in the "search-for-water-mode" until its thirst is completely satisfied - even if on its way to the water the most delicious fruits are located which could have been picked up as an opportunity to satisfy its hunger. While this model might work in a few selected conditions, it certainly will not produce a sensible and adaptive behavior in most circumstances. "The winner takes it all" in contrast can lead to an ineffective oscillation of behavior. In the competition of hunger and thirst the agent will drink only as long as the thirst motivation is slightly below the hunger motivation. At this moment, hunger takes control over behavior and directs movement toward the food resource. During travel time, both needs rise, and due to the dithering of behavior both needs will be unsatisfied in the long run.

An ethologically inspired solution to this problem is called a *time sharing mechanism*, first proposed by McFarland (1976). Time sharing describes the ability of an organism to allow low-priority goals to gain temporarily behavioral control even if a higher-priority goal is present. A computational model to achieve time-sharing consists of two mechanism: inhibition and fatigue. Inhibition occurs when an active motivational system inhibits competing motivations, leading to behavioral persistence. On the other hand, to prevent agents from mindlessly pursuing unreachable goals, a second mechanism, fatigue, is implemented. Fatigue is a dampening factor, which rises whenever specific activities are performed. That means, the longer an activity is performed without reaching its goal the stronger it is damped by fatigue, allowing other motivational systems to take a turn (for computational details of time sharing see Blumberg, 1994; Ludlow, 1980). Behavior controlled by the combination of inhibition and fatigue results in a hysteresis, where both

dithering and rigidity of behavior are avoided. The specific amount of behavioral persistence depends on the amount of inhibition and fatigue, resulting in some compromise between behavioral rigidity and behavioral oscillation.

Returning to the question of believability, a well-balanced mechanism of action selection, based on a set of motivations that is reasonable for the ECA's context, should enhance the impression of believability. While this topic is a standard problem each agent architect has to solve, there is only very few coverage in psychological research. Although there is a strong classical tradition about intraindividual motivational conflicts (approach-approach conflicts, approach-avoidance conflicts, etc.; Lewin, 1931), and several contemporary studies that deal with motivational conflict, most of them only investigate the consequences of those conflicts (e.g. an impaired well-being, e.g. Riediger & Freund, 2008), and not the underlying processes. Furthermore, the majority of these approaches only investigates one-shot decisions or cross-sectional data, and does not deal with the dynamic interplay of ongoing forces that compete for behavioral control. The action selection problem is not so much a problem of a single choice what to do in a concrete situation, but much more concerned with regulating, optimizing and balancing different drives over time. Studies investigating the processes how humans solve and self-regulate these motivational conflicts are rare (e.g. Kumashiro, Rusbult, & Finkel, 2008). In this case, design considerations of virtual agents point to a rather neglected field in psychological research, and could be an inspirational source for future studies.

Interpersonal Perception and Behavioral Synchrony

One main feature of ECAs is the ability to “recognize and respond to verbal and non-verbal input” and “to deal with conversational functions such as turn taking, feedback, and repair mechanisms” (Cassell, 2000, p. 70). Some ECAs have the ability to sense human users in the real environment by means of cameras or microphones. For example, the virtual agent

"MAX" (Kopp, Gesellensetter, Krämer, & Wachsmuth, 2005) perceives and tracks multiple persons standing in front of him with a camera. "REA" (Cassell, 2001) for instance interprets conversational pauses smaller than 500 ms such as the user wants some (non-verbal) feedback. The ability to process these sensory informations and to translate them into a meaningful and coherent communication, which, in turn, enhances believability, needs a detailed knowledge about human communication processes and interpersonal perception.

What can current psychological research contribute to the question of interpersonal perception? Unfortunately, most empirical work is based on aggregated measures of behavior and does not provide enough information about the processes, dynamics, or timing of communicative phenomena which would be necessary for a top-down implementation of these theories. Advanced models in current research of personality and interpersonal judgments, however, explicitly investigate the role of cues in interpersonal perception and, for example, seek to find which specific behavioral cues predict personality traits or which specific cues or cue-preference combinations predict interpersonal attraction. For example, Back, Schmukle, and Egloff (in press) investigated attraction at zero acquaintance in an extensive design with 2628 dyads. They could show how perceptible cues of the target affect attraction in general (e.g. pleasantness of voice, "babyfaceness", energy of body movements) and how preference similarities between perceiver and target predicted relational attraction. These findings can be of great value for the construction of ECAs, as many encounters in human-machine interactions are at zero acquaintance.

Other basic properties in non-verbal communication are the phenomena of synchronization and mimicry. Research concerning affective and behavioral synchronization demonstrates its importance for the adaptivity and quality of communicative processes. In the investigation of client-therapist dyads, Ramseyer and Tschacher (2008) found that synchrony (measured as the energy of body movements) predicted both the perceived quality of the

therapeutical relationship on a micro level (within each session) as well the therapeutical success on a macro level. Furthermore, in a non-clinical population, Chartrand and Bargh (1999) demonstrated that non-verbal mimicry (the "chameleon effect") served as a cause of interpersonal rapport and empathy, and consequently led to a smoother interaction. Comparable results could be found in opposite-sex dyads, where synchrony of behavioral patterns predicted interpersonal attraction (Grammer, Kruck, & Magnusson, 1998). Complementary, research about disordered communication in patients can clarify what ECAs should avoid as much as possible: Steimer-Krause, Krause, and Wagner (1990), for example, could show that a denial of affective synchronization (a behavior frequently found in schizophrenic patients) is an effective way to induce negative affect in the interaction partner. This and other findings could be a caveat for affective computing: Maybe no emotional display is sometimes better than a wrong emotion at the wrong time (see also Cassell, Bickmore, Campbell, Vilhjalmsson, & Yan, 1999, for the importance of timing in conversations).

To the authors knowledge only two studies by Bailenson and colleagues are present about the effect of agent-initiated mimicry: Agents that imitated the user's head position with a delay of 4 seconds where evaluated more persuasive and likable (Bailenson & Yee, 2005). This effect, however, only was true when participants did not explicitly detect the mimicry (Bailenson, Yee, Patel, & Beall, 2008). Although the implementation of mimicry in these studies was relatively simple, the study demonstrates the potential of this largely unutilized phenomenon for ECA's non-verbal communication.

The Zurich Model of Social Motivation

The purpose of this section is to reintroduce a model that might fulfill the needs of both worlds, the psychological world and the virtual agent world, providing both a psychologically sound theory of basic social motivations and the computational details that are needed to

implement the theory without too much of interpretation and reconstruction: the Zurich model of social motivation (referred to as *ZM* in the remaining paper; Bischof, 2001; Gubler & Bischof, 1991; Schneider, 2001). The ZM has some roots in Bowlby's attachment theory (Bowlby, 1980). As attachment theory is formulated in terms of control systems theory and information processing, it is supposed to be very suitable for simulations and several architectures have been developed based on attachment theory (Bischof, 1975; Horswill, 2008; Petters & Waters, 2010).

However, the ZM goes far beyond attachment theory concerning the broadness of social phenomena it covers. It claims to describe dynamic motivated behavior from a developmental, evolutionary, and systems theory perspective, not only in infants, but also in adolescents and adults. Furthermore, the ZM not only incorporates security seeking behavior, but postulates three phylogenetically old motivational systems: (a) the security system, (b) the arousal system and (c) the autonomy system (see Figure 1). The autonomy system furthermore is divided into three phylogenetically distinguishable motives: power, prestige and achievement. All of these motivational systems are modeled as feedback control cycles which compare an internal set point with an actual value that is perceived through specific detectors. The discrepancy between actual value and set point is the resulting motivational activation.

-- INSERT FIGURE 1 ABOUT HERE --

The security system is a formalization and expansion of the attachment system (Bowlby, 1980) and ensures that the contact to care givers does not get lost. For that purpose, the security system compares the actual value of felt security with an internal set point, called dependency. The ZM clearly defines the environmental cues that influence the actual value. In the case of security it depends on three perceivable cues of other objects: their familiarity, their distance, and their relevancy (a measure of the potency of the object to alter a situation;

it is highest for mature adults, medium for less potent siblings and low for inanimate objects like a teddy bear). These three input variables are combined multiplicatively, resulting in the actual value of felt security. Therefore maximum felt security is achieved, when a familiar and relevant conspecific (e.g. the mother) is nearby. If the actual value falls short of the set point the organism is in a state of security appetence, and thus proximity seeking behavior is initiated. However, if the actual value exceeds the set point, the organism feels an overabundance of security (i.e. security aversion) and surfeit behavior is triggered - a situation that is typically present in puberty. All other motivational control systems are modeled accordingly. The ZM not only describes these core control cycles, but also postulates interconnections between those systems as well as a coping system (not displayed in Figure 1), which gets in charge whenever a motivational activation does not get reduced for a longer time (Gubler & Bischof, 1993).

The scope of the article is not to elaborate the computational details of the model, as these are described in detail in the original publications. However, some examples are given to demonstrate the fine grained level of the model, which goes far beyond most other psychological models of motivation:

(a) action selection: The ZM proposes two types of hystereses to deal with the problem of action selection. Concerning the basic motivational systems a hysteresis with an implied hierarchy of motives is assumed (e.g., an aversion of arousal always has higher priority than an aversion of security; Gubler & Bischof, 1993). Concerning the autonomy system, a more sophisticated cusp catastrophe is proposed (and mathematically described) to model the dynamics of hierarchic encounters, where two opponents reciprocally build up an autonomy claim until the claim of one collapses and the hierarchy is stabilized again (Bischof, 1996).

(b) Interpersonal perception: The sensory inputs of all motivational systems are clearly defined. In an ECA with visual capabilities (e.g. MAX, Kopp et al., 2005), all sensors

theoretically could be implemented. Face detection algorithms in combination with a database that records the overall duration of interactions could form the familiarity sensor. Other face detection mechanisms that distinguish facial features could discriminate adults and children as a first approximation to detect the relevance. Other facial features for relevance/dominance could include a prominent chin, body postures, or gaze direction (Hall, Coats, & LeBeau, 2005).

Formulas for many other scenarios are provided as well: What happens if multiple familiar objects are nearby - are two moderately familiar persons better than one highly familiar? How is physical distance related to psychological distance? Is the set point always constant or can it be influenced?

Example Applications

In the following section, two implementations of the ZM will be shortly presented. For details of the implementations and the empirical results, please refer to the original articles referenced.

Dynamic emotional expressions: The varieties of smiling. Most implementations of emotions are based on some kind of appraisal theories, with the OCC model of emotions (Ortony et al., 1988) as the most prominent theory. This model computes a discrete emotion based on the current appraisal of the situation and internal factors, which then is expressed as a “fixed action pattern” by the agent (Gratch, 2008). There are some other approaches, like the WASABI architecture by Becker-Asano and Wachsmuth (2008) who implemented a continuous model of emotions based on the PAD space (Pleasure - Arousal - Dominance). However, when it comes to the expression of the emotion, the continuous emotional space still is mapped onto discrete emotional categories, and the agent expresses the prerecorded emotion with the highest likelihood. This reduces the believability of the agent because evolving appraisals of the situation or mixed emotions get lost (De Rosis et al., 2003; Gratch,

2008). Although the component process theory developed by Scherer and colleagues (Scherer, Schorr, & Johnstone, 2001) in principle describes the dynamic evolution of an emotional expression through the various stages of appraisal, this dynamic approach has not yet been implemented extensively (see, however, Paleari, Grizard, & Lisetti, 2007).

The ZM primarily pronounces the self-regulatory function of emotions, as emotions are only supposed to occur when the primary motivational reaction does not lead to a reduction of the motivational tension. In this case, some sort of coping has to take place, and emotions are supposed to be internal signals to activate and direct the coping system. Although the communicative aspect in this view is rather secondary, a specific mapping from internal motivational indices to facial expressions (action units) is made (Bischof, 1996). While some of these mappings show some similarities to Scherer's system (e.g. arousal appetence shares some attributes with Scherer's novelty check), others are hard to compare as the ZM has a stronger focus on dynamics of internal variables.

In one case the ZM makes particular detailed and explicite predictions: it proposes seven varieties of smiling (Bischof, 1996), discussing the fact that smiling is not always an indicator for joy or happiness. The ZM predicts that smiling occurs whenever the claim of autonomy is reduced (i.e., the first derivative of the autonomy claim is negative). Based on this assumption at least seven different types of smiling can be differentiated, corresponding to seven situations where the claim of autonomy is reduced. To display the facial expression the strength of the computed motivational indexes is directly mapped to the contraction of specific facial muscles. This also implies that the progression of the smile is not pre-scripted, but in fact dynamically responds to changing environmental inputs. Depending on the motivational system the relaxation of the autonomy claim can result in a trustful smile, smile of relief, embarrassed smile, anxious smile, surprised smile, superior smile, or inferior smile. In the case of the inferior smile, for example, the former claim of autonomy collapses when in

a hierarchical fight the inferior opponent withdraws. Each of the smiling types is clearly defined in terms of internal states (e.g., the set points of the security or autonomy system) and external signals that trigger the smile (e.g., the appearance of an unfamiliar person).

Borutta, Sosnowski, Zehetleitner, Bischof, and Kühnlenz (2009) implemented these seven types of smiling in animated avatar faces. In an evaluation study they could demonstrate that participants could classify the type of smiling in the resulting emotional video sequences significantly better than chance, supporting the plausibility of the proposed model of smiling and the underlying motivational dynamics. To the author's knowledge, this is one of the first implementations where continuous internal variables are dynamically mapped onto facial muscles (cf., however, Krumhuber, Manstead, & Kappas, 2007, for an evaluation of the dynamics of smiling). While the resulting "smiling head" is not yet a mature ECA at all (as sensory functions and conversational features are missing), the research demonstrates how the ZM can add believability and serves as a viable model for implementation in virtual agents.

Psychological assessment in virtual worlds: How do you treat your virtual spouse? In another project, several motivated agents were designed to interact in a multi-agent environment. The purpose of this study was to assess the participants' behavior towards his or her "virtual spouse" in the context of a larger study investigating romantic relationships (Schönbrodt & Asendorpf, in press). The simulation was presented as an online, interactive computer game where the participant could control one of the agents. In the story of the computer game, the user-controlled agent has a spouse. They are living in a community ("Simoland") together with some other motivated agents, and several scripted events take place during the 15 minutes of game play. The autonomous agents are controlled by a simplified version of the ZM and have a security system to seek contact to familiars (with associated behaviors like kissing, smooching, or talking about their relationship), an arousal

system to contact moderately familiar agents (e.g. dancing together, hearing music), and the non-social motives to regularly satisfy hunger and thirst. Due to the motivated nature of the agents, "life goes on" in Simoland, regardless whether the user interacts or not. The user can initiate more than 30 different behaviors or interactions, and from the resulting course of the game a diversity of game indices can be calculated for diagnostic purposes.

In this application, the motivated agents were not the aim of the investigation but rather provided a background for a new type of personality assessment. Although their conversational capabilities are very limited (e.g., activities and dialogs are displayed by symbols), participants really got involved into the game and developed an affection for the agents. In a pretest (n=19) we asked participants whether they experienced an emotional moment during the game and asked them to describe it in an open-ended question. The majority of participants did so, and answers like "When Lisa started to flirt with my husband, I really got jealous. I tried to distract her, so that she stopped flirting" supported the believability of the agents.

Concluding Remarks on the Zurich Model of Social Motivation

In comparison to other psychological models of motivation and emotion, the ZM has the unique feature of an explicit mathematical and computational base. Only very few other psychological theories have a comparable mathematical grounding (e.g. the "dynamics of action" approach, Atkinson & Birch, 1970). However, they usually are limited to a very narrow domain (e.g. achievement motivation).

In comparison to many other models of motivation developed from the agents community, an advantage of the ZM is its sound psychological foundation. It claims to incorporate all basic social motivations of humans, a claim which could be empirically supported (Schönbrodt, Unkelbach, & Spinath, 2009). Therefore, if the goal of a researcher is

to equip his or her agents with social motivations, the ZM can serve as an integrative and rather exhaustive model.

A potential drawback of the model, however, is that many publications about it are in German language only. Concerning the validity of the ZM, Gubler, Paffrath, and Bischof (1994) could predict participants' behavior in a space ship simulator by modeling their motivational dynamics with the ZM. In an English publication, Schönbrodt et al. (2009) could show that questionnaire scales assessing the set points of the motivational systems are significantly related to real-life outcomes; and, as described above, Borutta et al. (2009) could show that avatar smiles produced by the ZM could be correctly categorized (for other English publications, see Bischof, 1975; Borutta et al., 2009; Gubler & Bischof, 1991; Schneider, 2001). Given the complexity of the ZM, however, numerous future validation studies are needed to further explore the validity of the model.

Another potential problem is the behavioral output of the model (see “behavioral programs” in Figure 1). While the model is very detailed and explicit about the generation of motivations, it is rather limited concerning the precise behavioral programs that should be triggered by certain motivational states. Nonetheless, from our point of view the model seems to be a good starting point for the implementation of social motives into autonomous agents.

Conclusion

In this article it is argued that psychological models of motivation, emotion, interpersonal perception, and non-verbal communication can enrich virtual agents by enhancing their believability. But it is not only the virtual agents community that can benefit from psychological theories - the embodiment of psychological theories into virtual agents can be a fruitful step in the process of theory construction and testing in psychology as well. As Karl Grammer states, “much [about psychology] can be learned from reverse engineering” (Schönbrodt, 2007), and many possibilities to improve psychological theories become

apparent when one tries to implement them. We hope that this article might serve as a starting point for interested psychologists as well as an inspiration for the development of psychologically more believable agents.

References

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action.* Oxford, England: John Wiley.

Back, M., Schmukle, S. C., & Egloff, B. (in press). **A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance.** ***European Journal of Personality.***

Bailenson, J., & Yee, N. (2005). Digital chameleons. *Psychological Science*, *16,* 814-819.

Bailenson, J., Yee, N., Patel, K., & Beall, A. (2008). Detecting digital chameleons. *Computers in Human Behavior*, *24,* 66-87.

Bates, J. (1994). The role of emotion in believable characters. *Communications of the ACM*, *37,* 122-125.

Becker-Asano, C., & Wachsmuth, I. (2008). Affect simulation with primary and secondary emotions. *Lecture Notes in Computer Science*, *5208,* 15-28.

Beutler, L., & Harwood, T. (2004). Virtual reality in psychotherapy training. *Journal of Clinical Psychology*, *60,* 317-330.

Bischof, N. (2001). *Das Rätsel Ödipus. Die biologischen Wurzeln des Urkonflikts von Intimität und Autonomie. [The riddle of Oedipus. The biological roots of the core conflict between intimacy and autonomy].* München: Piper.

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

Bischof, N. (1996). Untersuchungen zur Systemanalyse der sozialen Motivation IV: Die Spielarten des Lächelns und das Problem der motivationalen Sollwertanpassung [The varieties of smiling and the problem of motivational adjustment]. *Zeitschrift für Psychologie*, *204,* 1-40.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social

psychology. *Psychological Inquiry*, *13,* 103-124.

Blumberg, B. (1994). *Action-selection in Hamsterdam: Lessons from ethology*. Paper presented at the Third Conference on the Simulation of Adaptive Behavior, Cambridge, MA.

Borutta, I., Sosnowski, S., Zehetleitner, M., Bischof, N., & Kühnlenz, K. (2009). *Generating artificial smile variations based on a psychological system-theoretic approach*. Paper presented at the 18th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man), Toyama, Japan.

Bowlby, J. (1980). *Attachment and loss.* New York, NY, US: Basic Books.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., & Yan, H. (1999). Conversation as a system framework: Designing embodied conversational agents. In J. Cassell, E. Churchill, S. Prevost, & J. Sullivan (Eds.), *Embodied conversational agents*. Cambridge: MIT Press.

Cassell, J., Churchill, E., Prevost, S., & Sullivan, J. (2000). *Embodied conversational agents.* Cambridge: MIT Press.

Cassell, J. (2000). More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM*, *43,* 70-78.

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Magazine*, *22,* 67-83.

Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76,* 893-910.

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & Carolis, B. (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, *1-2,* 81-118.

Fridlund, A. J. (1991). Evolution and facial action in reflex, social motive, and paralanguage.

Biological Psychology, *32,* 3-100.

Grammer, K., Kruck, K., & Magnusson, M. (1998). The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal Behavior*, *22,* 3-29.

Gratch, J. (2008). True emotion vs. social intentions in nonverbal communication: Towards a synthesis for embodied conversational agents. *Lecture Notes in Computer Science*, *4930,* 181-197.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5,* 269-306.

Gubler, H., & Bischof, N. (1993). Untersuchungen zur Systemanalyse der sozialen Motivation II: Computerspiele als Werkzeug der motivationspsychologischen Grundlagenforschung [Computer games as a tool for basic research on motivation]. *Zeitschrift für Psychologie*, *201,* 287-315.

Gubler, H., & Bischof, N. (1991). A systems theory perspective. In M. E. Lamb & H. Keller (Eds.), *Infant development: Perspectives from German-speaking countries* (pp. 35-66). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Gubler, H., Paffrath, M., & Bischof, N. (1994). Untersuchungen zur Systemanalyse der sozialen Motivation III: Eine Ästimationsstudie zur Sicherheits- und Erregungsregulation während der Adoleszenz [An estimation study of security and arousal regulation during adolescence]. *Zeitschrift für Psychologie, 202,* 95-132.

Hall, J., Coats, E., & LeBeau, L. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, *131,* 898-924.

Horswill, I. (2008). *Attachment and cognitive architecture*. Paper presented at the 2008 AAAI Spring Symposium.

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human–computer interaction.

International Journal of Human-Computer Studies, *59,* 1-32.

Kopp, S., Gesellensetter, L., Krämer, N., & Wachsmuth, I. (2005). A conversational agent as museum guide-Design and evaluation of a real-world application. *Lecture Notes in Computer Science*, *3661,* 329-343.

Krämer, N., Bente, G., Eschenburg, F., & Troitzsch, H. (2009). Embodied conversational agents. *Social Psychology*, *1,* 26-36.

Krämer, N., Iurgel, I., & Bente, G. (2005). *Emotion and motivation in embodied conversational agents*. Paper presented at the Proceedings of the Symposium "Agents that want and like", Artificial Intelligence and the Simulation of Behavior (AISB) 2005, Hatfield.

Krämer, N. C. (2008). *Soziale Wirkungen virtueller Helfer. Gestaltung und Evaluation von Mensch-Computer-Interaktion [Social effects of virtual assistants].* Stuttgart: Kohlhammer.

Krumhuber, E., Manstead, A., & Kappas, A. (2007). Temporal aspects of facial displays in person and expression perception: The effects of smile dynamics, head-tilt, and gender. *Journal of Nonverbal Behavior*, *31,* 39-56.

Kumashiro, M., Rusbult, C., & Finkel, E. (2008). Navigating personal and relational concerns: The quest for equilibrium. *Journal of Personality and Social Psychology*, *95,* 94-110. doi: 10.1037/0022-3514.95.1.94

Lewin, K. (1931). Environmental forces in child behavior and development. In C. Murchison (Ed.), *Handbook of child psychology* (pp. 94-127). Worcester, Mass: Clark University Press.

Ludlow, A. (1980). The evolution and simulation of a decision maker. In F. Toates & T. Halliday (Eds.), *Analysis of Motivational Processes.* London: Academic Press.

McFarland, D. J. (1976). Form and function in the temporal organisation of behavior. In P.

Bateson & R. Hinde (Eds.), *Growing points of ethology (pp. 55-93)*. Cambridge University Press.

Ortony, A. (2003). On making believable emotional agents believable. In R. Trappl, P. Petta, & S. Payr (Eds.), *Emotions in humans and artifacts* (pp. 189-212).

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions.* New York: Cambridge University Press.

Paleari, M., Grizard, A., & Lisetti, C. (2007). Adapting psychologically grounded facial emotional expressions to different anthropomorphic embodiment platforms. In *Proceedings of the 20th Florida Artificial Intelligence Research Society FLAIRS 2007 Annual Conference on Artificial Intelligence, Florida.*

Petters, D., & Waters, E. (2010). *AI, attachment theory and simulating secure base behaviour: Dr. Bowlby meet the Reverend Bayes.* Manuscript submitted for publication.

Ramseyer, F., & Tschacher, W. (2008). Synchrony in dyadic psychotherapy sessions. In S. Vrobel, O. E. Rössler, & T. Marks-Tarlow (Eds.), *Simultaneity: Temporal structures and observer perspectives* (pp. 329-347). Singapore: World Scientific.

Redin, J. (2007). A brief history of mechanical calculators. Retrieved from http://www.xnumber.com/xnumber/mechanical1.htm

Riediger, M., & Freund, A. (2008). Me against myself: Motivational conflicts and emotional development in adulthood. *Psychology and Aging*, *23,* 479-494.

Rose, R., Scheutz, M., & Schermerhorn, P. (2010). Towards a conceptual and methodological framework for determining robot believability. *Interaction Studies, 11,* 314-335.

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research.* Oxford University Press, USA.

Schneider, M. (2001). Systems theory of motivational development. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences.* Oxford:

Elsevier.

Schönbrodt, F. D. (2007). Interview mit Prof. Dr. Karl Grammer. *Zeitschrift für Medienpsychologie*, *4,* 160-161.

Schönbrodt, F. D., & Asendorpf, J. B. (in press). Virtual social environments as a tool for psychological assessment: Dynamics of interaction with a virtual spouse. *Psychological Assessment*.

Schönbrodt, F. D., Unkelbach, S., & Spinath, F. M. (2009). Broad motives in short scales: a questionnaire for the Zurich Model of social motivation. *European Journal of Psychological Assessment*, *25,* 141-149.

Steimer-Krause, E., Krause, R., & Wagner, G. (1990). Interaction regulations used by schizophrenic and psychosomatic patients: Studies on facial behavior in dyadic interactions. *Psychiatry: Journal for the Study of Interpersonal Processes*, *53,* 209-228.

Wehrle, T. (1998). *Motivations behind modeling emotional agents: Whose emotion does your robot have*. Paper presented at the Grounding Emotions in Adaptive Systems. Zurich: 5th International Conference of the Society for Adaptive Behavior Workshop Notes (SAB'98), Zurich, Switzerland.

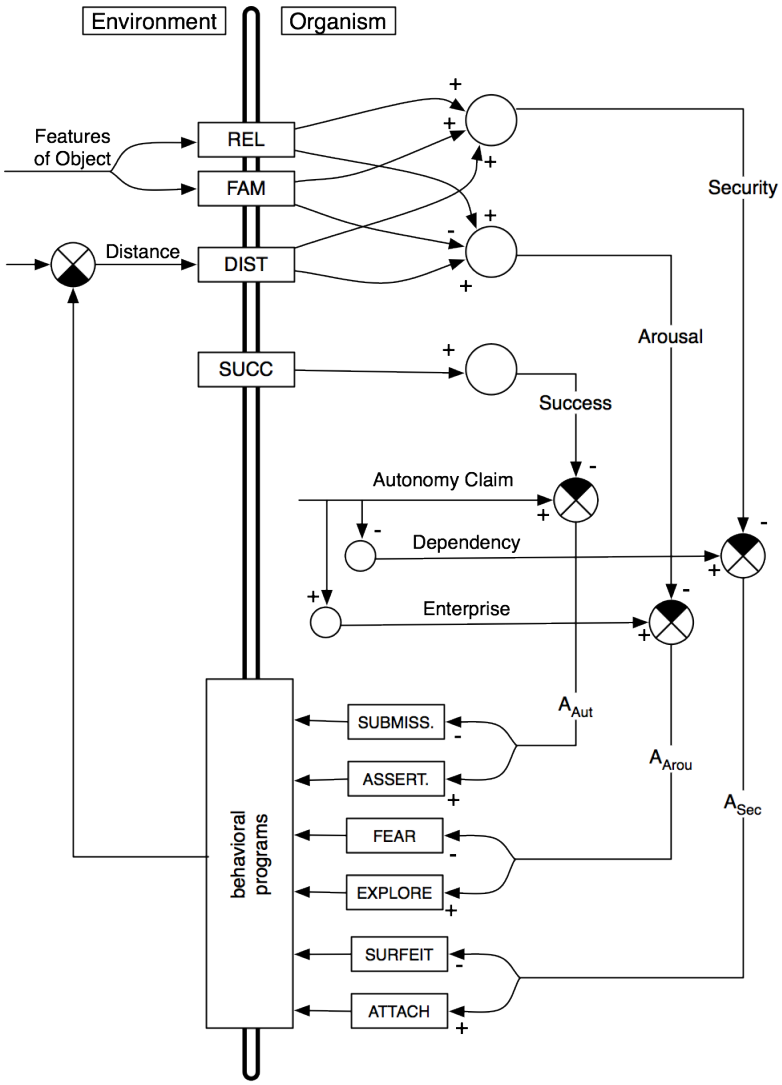

Figure 1. A simplified version of the Zurich model of social motivation (adapted from Bischof, 1993). REL= relevance, FAM = familiarity, DIST = distance, SUCC = success, A_{Aut}= activation of autonomy system, A_{Arou}= activation of arousal system, A_{Sec}= activation of security system.