Unpacking the Distinct Roles of Mimicry and Synchrony in Emotion Regulation and Bonding

Ethel Pruss
Department of Clinical Psychology
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
e.pruss@vu.nl
ORCID: 0000-0003-4281-5577

Maryam Alimardani
Department of Computer Science
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
m.alimardani@vu.nl
ORCID: 0000-0003-3077-7657

Sascha Struijs
Department of Clinical Psychology
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
s.y.struijs@vu.nl
ORCID: 0000-0002-2668-5410

Sander L. Koole
Department of Clinical Psychology
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
s.l.koole@vu.nl
ORCID: 0000-0002-3422-223X

***Abstract*—Virtual humans (VHs) provide a powerful platform for experimentally probing how nonverbal coordination shapes emotional and relational processes. This paper outlines a planned study with a twofold aim. First, it seeks to clarify how timing and form of nonverbal coordination contribute to emotional coregulation and bonding in human interaction, leveraging VHs for the precise orthogonal manipulation of these processes. Second, it aims to inform the development of VHs with natural and effective nonverbal responsiveness, supporting applied contexts such as mental health interventions and VH companions. The planned study has a 2×2 factorial design in which VH behavior will be manipulated along two dimensions: movement timing (synchrony) and movement form (mimicry). Participants will engage in semi-structured, emotionally themed conversations with an LLM-powered VH acting in a supportive coach role. During the participant's speaking phases, the VH's nonverbal responses will vary according to the experimental condition, while verbal responses and baseline behaviors remain naturalistic. Outcomes will be assessed at three levels: self-reported emotion regulation and bonding outcomes, and behavioral indices of coordination derived from video data.**

***Index Terms*—Virtual humans, interpersonal coordination, mimicry, synchrony, nonverbal behavior, emotional coregulation, bonding.**

I. Introduction

Virtual humans (distinct from avatars) are human-like AI-driven agents that exhibit their own agency and behaviors. They offer controlled settings for studying how nonverbal dynamics affect social outcomes because researchers can program movements with precise timing and spatial properties and measure participant responses [1]. Emotional coregulation can be defined as the interpersonal sharing and joint regulation of affective states [2]. Along with the development of bonding and rapport, it is thought to rely on subtle forms of nonverbal coordination [3]–[6]. However, little is known about how specific parameters, for example, precise temporal coordination (synchrony) versus matching form (mimicry), shape these processes. Beyond their experimental utility, virtual humans also have applied contexts where emotional coregulation may be beneficial, such as companions or mental health aids, making it important to understand which forms of responsive nonverbal behavior are most effective. The following section reviews empirical work on nonverbal mimicry and synchrony in virtual-human interactions and highlights the need to distinguish these constructs.

II. Related Work

A. Mimicry in Virtual Human Interactions

Behavioral mimicry involves replicating another's gestures, postures, or expressions, typically after a short delay [3], [7]. Behavioral mimicry has been found to foster positive social outcomes such as increased prosocial behavior, rapport and social bonding [8]–[11]. While evidence in virtual human interactions is more scarce, similar effects have been found. For example, [12] found that mimicry strengthened rapport and trust: for example, when an agent reproduced participants' smiles during storytelling, participants felt more liking and trust toward the agent. Another study had a virtual coach mimic head and body movements with random 0.5–4 second delays; both schizophrenia patients and controls reported greater comfort and showed higher behavioral synchronization with the mimicking agent than with a non-mimicking agent [13]. These findings support the "social glue" hypothesis that being mimicked fosters positive social outcomes [10]. However, results are mixed: some virtual agents elicited no increase in rapport or trust when they mirrored users' movements [1], and other experiments found mimicry improved trust only in certain game contexts but not others [14]. Null effects may stem from differences in experimental tasks, participant awareness, or cultural factors [1], and thus mimicry's influence on emotional coregulation and bonding remains context-dependent.

B. Synchrony in Virtual Human Interactions

Synchrony refers to aligning the timing of actions without necessarily matching the specific movements; it often feels like moving "in sync" with another [3], [7]. Synchrony is often defined as simultaneous motor behavior [15], whereas mimicry refers to copying behavior after a 3–5 second lag [16], [17]. Similarly to mimicry, synchrony is associated with positive interpersonal outcomes, such as increased rapport and closeness [15], [18]–[20]. Evidence on synchrony in virtual humans is sparse. One immersive VR study representing participants as virtual humans manipulated whether co-participant movements were synchronous or not; participants reported greater social closeness in the synchrony condition [21]. The authors emphasized that traditional synchrony experiments often rely on explicit instructions or shared rhythms (e.g., dancing to music), confounding synchrony with joint intention or mimicry [21]. Another VR "mirror game" study programmed a character to loosely couple its movements to the user's in a joint-imitation task; participants rated the fluidity and influence of the interaction higher when coupling was present and could not distinguish a loosely coupled agent from a human partner [22]. These studies suggest synchrony can enhance rapport and perceived connectedness, yet they also highlight how infrequent and varied synchrony manipulations are in virtual-human research.

C. Challenges in Disentangling Mimicry and Synchrony

In human interactions, mimicry and synchrony are difficult to separate because they naturally occur together. Virtual human experiments enable the manipulation of form and timing independently, but the two factors are still often not clearly separable in previous studies. For instance, a VR experiment that compares 1–3 second delays treat a shorter delay as more synchronous, but the agent also replicates the user's movement form for all delay conditions [1]. Conversely, mimicry manipulations sometimes involve immediate imitation, blending form matching and temporal alignment [13]. Timing conventions are also inconsistent across the literature, both for human-human and human-agent interactions. While a delay of 3–5 seconds is often treated as the standard for mimicry [16], [17], some studies allow shorter lags that could reasonably fall into the domain of synchrony. Similarly, synchrony is often operationalized as perfectly simultaneous behavior [3], [7], but other studies allow several seconds of lag, making it difficult to distinguish from delayed mimicry [15]. As a result, it is unclear whether the psychological and social effects that have been observed arise from matching the exact movement, synchronizing timing, or both. For example, emotional coregulation may hinge on subtle timing cues rather than mirror-like copying, and bonding may be sensitive to whether the agent follows the user's motions or merely moves in the same tempo. Existing evidence thus underscores the need for experiments that orthogonally manipulate form and timing to isolate their contributions to emotional coregulation and bonding with virtual humans.

D. Research Goals

To address the challenges outlined above in disentangling mimicry and synchrony, and the lack of empirical research on how nonverbal coordination influences emotional coregulation, this paper introduces the methods for a planned study. The planned study tackles both gaps through a controlled experimental design that manipulates mimicry and synchrony in virtual-human interactions using a 2×2 factorial design. By varying movement form and timing independently during emotionally themed conversations with a supportive virtual coach, we aim to disentangle their respective contributions to emotional coregulation and bonding, which will be treated as outcome variables. Our approach has a twofold purpose. First, it leverages virtual humans to clarify how fundamental mechanisms of nonverbal alignment shape human emotional and social processes. Second, it works toward the development of virtual humans with natural and effective nonverbal responses, a prerequisite for applied contexts such as mental health support and companion technologies. In doing so, we provide a methodological approach that both advances basic research on nonverbal coordination and supports its translation into applied human–agent interaction.

III. Methods

A. Participants

We will recruit eighty English-speaking university students. The planned sample size is based on detecting small–to–medium effects (Cohen's $d \approx .35$; $f \approx .175$) with an additional margin of twenty percent to accommodate exclusions and model complexity. Individuals who report hearing or visual impairments will be excluded. All participants will provide informed consent prior to any study procedures, and ethical approval will be obtained before recruitment begins.

B. Design and Overview

The study will adopt a within-subject 2×2 factorial design that will manipulate the virtual human's (VH) nonverbal behavior along two dimensions: Mimicry (present vs. absent) and Synchrony (present vs. absent). Each participant will complete four conversation blocks, corresponding to the four combinations of the factors, with each block lasting approximately five minutes. The order of conditions will be counterbalanced across participants, and conversation topics will be rotated to reduce content-specific effects (see Figure 1 for an overview).

C. Task and Virtual Human Behavior

Across the session, participants will engage in a 5-minute conversation with a virtual human on emotional topics related to the student population, such as procrastination and academic stress, loneliness and difficulties connecting with others, feeling down or emotionally fatigued, and uncertainty about future study or career choices. The topic will be predefined as one of these options (counterbalanced across conditions), and both the virtual human and participant will be informed of the topic at the start of the interaction. The virtual human will act as a coach. Speaking and listening phases will alternate and will

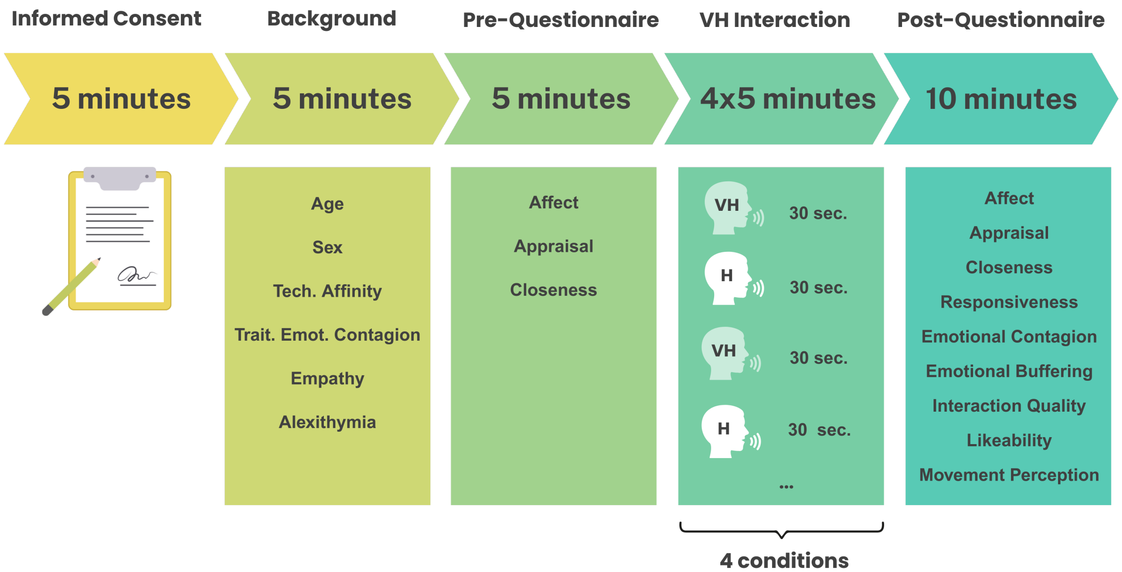

Fig. 1. Experiment Protocol and Timeline.

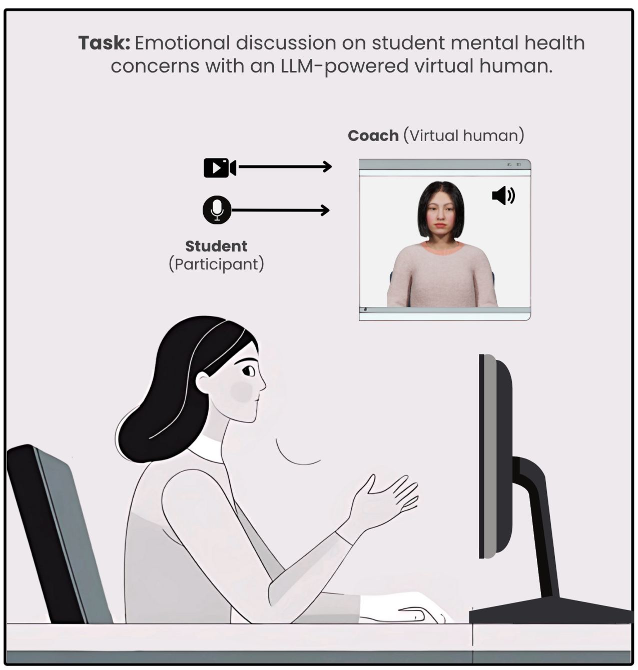

Fig. 2. Illustration of Experiment Setup.

be signaled by a short auditory cue so that both the participant and the VH can shift roles smoothly. During the VH's speaking phases, the agent will produce LLM-driven verbal responses accompanied by naturalistic co-speech gestures, while baseline realism behaviors (e.g., blinking, breathing, and subtle eye movements) will remain active throughout the interaction to maintain a realistic appearance. See Figure 2 for an illustration of the experiment setup.

D. Nonverbal Manipulation

The experimental manipulation will be implemented specifically during the VH's listening phases (i.e., while the participant is speaking). Two anatomical regions will be targeted for movement control: head and body, so that the system can selectively match or diverge from the participant's movement region. In the *No Mimicry / No Synchrony* condition, the VH will exhibit random movements sampled from the frequency and range distributions of pilot experiments. In the *Mimicry / No Synchrony* condition, the VH will reproduce the detected movement but after a delay of 3-5 seconds, thereby matching form without temporal coordination. In the *Synchrony / No Mimicry* condition, the VH will move synchronously with the participant, initiating movements within zero to two seconds of the detected movement, but will select a random gesture from the opposing region to avoid copying the form of the movement. Finally, in the *Mimicry + Synchrony* condition, the VH will reproduce the detected movement within zero to two seconds, aligning both form and timing. These settings will preserve a consistent conversational baseline while isolating the distinct contributions of the alignment in form and timing of responsive nonverbal behavior.

E. System Architecture

The VH system will be built in Unreal Engine [23] with MetaHuman [24] to generate a realistic human-like avatar, in this case, a young adult female (see Figure 2). Audio input will be captured and transcribed via speech-to-text, enabling GPT-4o to generate verbal responses together with gesture specifications. GPT-4o will be prompted to primarily validate and summarize the user's feelings in its responses, with the aim of making the participant feel supported. These responses will be converted to audio and viseme sequences using Azure TTS [25], [26], allowing for accurate lip-syncing and expressive speech delivery. Participant movements will be detected through webcam-based posture and gesture recognition using MediaPipe [27], informing the VH's responsive behaviors. Baseline motions such as breathing, blinking, and subtle eye movements will remain active in all conditions to sustain realism. Conversations will alternate between speaking and listening phases, signaled by a brief auditory cue. In speaking phases, the VH will deliver GPT-4o–driven responses with context-appropriate co-speech gestures and synchronized speech. In listening phases, the VH will perform nonverbal behaviors according to the assigned experimental condition.

The detected gestures will include head shakes, nods, and tilts (left/right) in the head region, as well as torso leans (in/back), shoulder shrugs, and hand-to-chin or hand-to-chest touches in the body region. These gestures will be motion-captured to ensure smooth and natural animation when displayed by the virtual human.

F. Measures

Outcomes will be assessed at three levels (see Figure 1 for an overview). For emotion regulation, participants will complete brief affect ratings immediately before and after each block using the PANAS short form and will provide appraisals of the interaction using items adapted from the Geneva appraisal framework [28]. For bonding, participants will report perceived closeness using the Inclusion of Other in the Self (IOS) scale [29], as well as likeability [30], perceived partner responsiveness, and overall interaction quality after each condition. To capture behavior, video-based pose tracking will yield movement time series from which we will compute indices of mimicry and synchrony (e.g., windowed cross-correlations, gesture-match counts, and lag distributions) aligned to speaking and listening phases. A brief manipulation check will probe the extent to which the VH's movement was perceived to be similar to one's own movement, and the extent to which the VH appeared to be moving in general during the interaction [31]. In addition, background questionnaires will index individual differences relevant to interpersonal emotion processing, including alexithymia (TAS-20) [32], empathy (IRI) [33], and susceptibility to emotional contagion [34].

G. Procedure

After providing informed consent, participants will complete background questionnaires and receive instructions about the structure of the conversational task. They will then engage in four conversation blocks, one per condition, with topics counterbalanced across blocks to minimize topic–condition confounds. Before and after each block, participants will complete questionnaires that capture emotion regulation and bonding outcomes (see Figure 1 for an overview). The full session is expected to last approximately forty-five minutes.

H. Data Processing

Video streams will be processed to extract pose trajectories for the defined movement regions. These trajectories will be segmented by role (participant speaking vs. VH speaking) and by condition, then preprocessed to ensure temporal alignment with the auditory cues marking turns. From these time series, we will derive complementary metrics of nonverbal coordination. Synchrony will be quantified using windowed cross-correlations with near zero lag to estimate temporal alignment, whereas mimicry will be indexed by form similarity and gesture matches within lag windows characteristic of delayed copying (3-5 seconds). Questionnaire scales will be scored per their published keys.

I. Statistical Analysis

Primary analyses will use linear mixed-effects models with Mimicry, Synchrony, and their interaction as predictors for emotional coregulation and bonding outcomes (such as affect, appraisal, closeness and responsiveness), while controlling for topic, condition order and baseline values. Random intercepts will be included for participants, with random slopes added when possible to capture individual differences. To test whether participants' nonverbal responses during the VH's speaking phases explain the link between the manipulations and outcomes such as bonding and emotion regulation, we will use multilevel mediation. Results will be reported as standardized estimates with 95% confidence intervals, and false discovery rate procedures will be used to adjust for multiple comparisons.

IV. Conclusion

This paper outlines a methodological framework for disentangling the roles of mimicry and synchrony in human–agent interaction. By leveraging virtual humans to independently manipulate nonverbal coordination in terms of movement timing and form during emotional conversations, the planned study seeks to clarify how these two factors independently contribute to emotional coregulation and bonding. Once completed, the study is expected to make two main contributions. First, it will provide empirical evidence on the distinct and joint effects of mimicry and synchrony, advancing theoretical understanding of how nonverbal coordination supports emotional and relational processes in humans. Second, it will develop practical design principles for creating virtual humans with natural and effective nonverbal behavior, advancing their use in mental health support and companion technologies.

References

[1] J. Hale and A. Hamilton, "Testing the relationship between mimicry, trust and rapport in virtual reality conversations," *Scientific Reports*, vol. 6, p. 35295, 2016.

[2] E. A. Butler and A. K. Randall, "Emotional coregulation in close relationships," *Emotion Review*, vol. 5, no. 2, pp. 202–210, 2013.

[3] J. L. Lakin, "Behavioral mimicry and interpersonal synchrony," in *Nonverbal Communication*, J. A. Hall and M. L. Knapp, Eds. Berlin, Boston: De Gruyter Mouton, 2013, pp. 539–576.

[4] S. L. Koole, D. Atzil-Slonim, E. Butler, S. Dikker, W. Tschacher, and T. Wilderjans, "In sync with your shrink: Grounding psychotherapy in interpersonal synchrony," in *Applications of Social Psychology*. Routledge, 2020, pp. 161–184.

[5] S. L. Koole and W. Tschacher, "Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance," *Frontiers in psychology*, vol. 7, p. 862, 2016.

[6] S. C. F. Hendrikse, J. Treur, T. F. Wilderjans, S. Dikker, and S. L. Koole, "On becoming in sync with yourself and others: An adaptive agent model for how persons connect by detecting intrapersonal and interpersonal synchrony," *Human-Centric Intelligent Systems*, vol. 3, no. 2, pp. 123–146, 2023.

[7] M. J. Hove and J. L. Risen, "It's all in the timing: Interpersonal synchrony increases affiliation," *Social Cognition*, vol. 27, no. 6, pp. 949–960, 2009.

[8] R. B. van Baaren, R. Holland, K. Kawakami, and A. van Knippenberg, "Mimicry and prosocial behavior," *Psychological Science*, vol. 15, pp. 71 – 74, 2004.

[9] M. Stel and R. Vonk, "Mimicry in social interaction: benefits for mimickers, mimickees, and their interaction." *British journal of psychology*, vol. 101 Pt 2, pp. 311–23, 2010.

[10] W. Kulesza, P. Muniak, M. Czekiel, S. Bedyńska, and A. Cislak, "Mimicry or responsiveness? verifying the mimicry-as-a-social-glue hypothesis," *Polish Psychological Bulletin*, 2023.

[11] N. Guéguen and A. Martin, "Incidental similarity facilitates behavioral mimicry," *Social Psychology*, vol. 40, pp. 88–92, 2009.

[12] E. L. Fasya, E. van den Bos, D. K. J. Heylen, and M. E. Kret, "Smile mimicry smoothens human–virtual human interactions," *International Journal of Human-Computer Studies*, vol. 183, p. 103182, 2024.

[13] S. Raffard, R. N. Salesse, C. Bortolon, B. G. Bardy, J. Henriques, L. Marin, D. Stricker, and D. Capdevielle, "Using mimicry of body movements by a virtual agent to increase synchronization behavior and rapport in individuals with schizophrenia," *Scientific Reports*, vol. 8, no. 1, p. 17356, 2018.

[14] F. M. F. Verberne, J. Ham, A. Ponnada, and C. J. H. Midden, "Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust," in *Persuasive Technology*, ser. Lecture Notes in Computer Science, S. Berkovsky and J. Freyne, Eds. Berlin, Heidelberg: Springer, 2013, vol. 7822.

[15] K. Fujiwara, M. Kimura, and I. Daibo, "Rhythmic features of movement synchrony for bonding individuals in dyadic interaction," *Journal of Nonverbal Behavior*, vol. 44, pp. 173–193, 2019.

[16] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception–behavior link and social interaction," *Journal of Personality and Social Psychology*, vol. 76, no. 6, pp. 893–910, 1999.

[17] B. Rauchbauer and M.-H. Grosbras, "Developmental trajectory of interpersonal motor alignment: Positive social effects and link to social cognition," *Neuroscience and Biobehavioral Reviews*, vol. 118, pp. 411–425, 2020.

[18] D. Lakens and M. Stel, "If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity," *Social Cognition*, vol. 29, pp. 1–14, 2011.

[19] D. Lakens, "Movement synchrony and perceived entitativity," *Journal of Experimental Social Psychology*, vol. 46, pp. 701–708, 2010.

[20] P. Reddish, R. Fischer, and J. Bulbulia, "Let's dance together: Synchrony, shared intentionality and cooperation," *PloS One*, vol. 8, no. 8, p. e71182, 2013.

[21] B. Tarr, M. Slater, and E. Cohen, "Synchrony and social connection in immersive virtual reality," *Scientific Reports*, vol. 8, p. 3693, 2018.

[22] J. Llobera, V. Jacquat, C. Calabrese, and C. Charbonnier, "Playing the mirror game in virtual reality with an autonomous character," *Scientific Reports*, vol. 12, no. 1, p. 21329, 2022.

[23] E. Games, "Unreal engine," https://www.unrealengine.com, 2025, accessed: 2025-09-29.

[24] ——, "Metahuman," https://www.unrealengine.com/metahuman, 2025, accessed: 2025-09-29.

[25] Microsoft, "Azure speech to text (speech service)," https://learn.microsoft.com/azure/ai-services/speech-service/speech-to-text, 2025, accessed: 2025-09-29.

[26] ——, "Azure text to speech (speech service)," https://learn.microsoft.com/azure/ai-services/speech-service/text-to-speech, 2025, accessed: 2025-09-29.

[27] Google, "Mediapipe solutions guide," https://ai.google.dev/edge/mediapipe/solutions/guide, 2025, last updated: 2025-09-09, accessed: 2025-09-29.

[28] K. R. Scherer, "Appraisal considered as a process of multi-level sequential checking," in *Appraisal Processes in Emotion: Theory, Methods, Research*, K. R. Scherer, A. Schorr, and T. Johnstone, Eds. New York and Oxford: Oxford University Press, 2001, pp. 92–120.

[29] A. Aron, E. N. Aron, and D. Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness," *Personality and Social Psychology Bulletin*, vol. 18, no. 3, pp. 242–251, 1992.

[30] A. Cerekovic, O. Aran, and D. Gatica-Perez, "How do you like your virtual agent?: Human-agent interaction experience through nonverbal features and personality traits," in *Intelligent Virtual Agents*, 2014, pp. 1–13.

[31] D. Jáuregui, T. Giraud, B. Isableu, and J.-C. Martin, "Design and evaluation of postural interactions between users and a listening virtual agent during a simulated job interview," *Computer Animation and Virtual Worlds*, vol. 32, no. 3–4, 2021.

[32] R. M. Bagby, J. D. A. Parker, and G. J. Taylor, "The twenty-item toronto alexithymia scale–i. item selection and cross-validation of the factor structure," *Journal of Psychosomatic Research*, vol. 38, no. 1, pp. 23–32, 1994.

[33] M. H. Davis, "A multidimensional approach to individual differences in empathy," *JSAS Catalog of Selected Documents in Psychology*, vol. 10, p. 85, 1980.

[34] R. W. Doherty, "The emotional contagion scale: A measure of individual differences," *Journal of Nonverbal Behavior*, vol. 21, no. 2, pp. 131–154, 1997.