4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

Pia Tikka*
Robert G. McNamara
Enactive Virtuality Lab
Baltic Film, Media and Arts School
Tallinn University
Tallinn, Estonia

Lynda Joy Gerry
Auckland Bioengineering Institute
University of Auckland
Auckland, New Zealand

Ilkka Kosunen
imec, Utrecht, The Netherlands

Mauri Kaipainen
Department of Cognitive Science
University of Helsinki
Helsinki, Finland

Designing Enactive Co-Presence: Second-Order Simulation of Empathy for Artificial Humans in Narrative Contexts

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

Abstract

Human interactions with human-like Computer-Generated Others (CGOs) have continued to become more prevalent in the rush to technologize societies. As such, the need for close analysis of mediative dynamics between humans and CGOs has become compelling. Relying on the theoretical framework of embodied enactive cognition, this paper will discuss designing human *empathetic experiences* of co-presence with CGOs. We propose a two-level idea of *enactive simulation*: the assumption of the first-order simulation *(human simulating other)* is built on the holistic assumption that humans understand others thanks to shared physical embodiment and enactive situatedness, while *second-order simulation* (designer simulating *human simulating other)* is the cognitive-affective means of the designer of human–machine encounters to apply their own first-order experiences to simulate the simulation of the Experiencer. Our key assumption is that context is the defining component of co-presence, thereby allowing experimental situations of co-presence to be contextualized by means of mediated narratives. The case study experiment, *Booth,* designed to model affective-cognitive experiences of co-presence at encounters of human Experiencers with two artificial asylum seekers, allowed us to identify different factors of the dynamics influencing the experience of empathy and the ensuing decisions concerning asylum, in particular, the contextualizing narrative.

1 Introduction

Can artificial computer-generated characters evoke a similar sense of co-presence as humans? How and to what extent can the experience of co-presence be guided? What is unique about them when it comes to modulating empathy? What role does their design play in their affordances for influencing moral perspective? These questions, among others, relate to the increasing application of artificial characters, referred to as Computer-Generated Others (CGO), not only in the field of entertainment and the game industry but also in the service of society. Our focus is on the design processes of co-presence between humans and CGOs in shared narrative contexts.

To set the ground, we draw from the *theory of enactive mind* in order to discuss affective-cognitive processes involved in designing *enactive narrative*

Presence, *Vol. 30, Winter 2021*, 149–166
https://doi.org/10.1162/PRES_a_00378
© *2023 by the Massachusetts Institute of Technology*

*Correspondence to pia.tikka@gmail.com.

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

systems. To facilitate this discussion, we first describe the *triadic model of enactive co-presence* between humans (later, Experiencer) and CGOs in particular situations (later, Context). Our approach is to interpret the determinant context as a temporally extended narrative, which can be mediated by means of cinema, virtual reality experiences, games, and so on. This allows us to further move the definition of the method of *enactive second-order simulation* with the objective of describing the creative process of designing contextually meaningful co-presence between Experiencer and virtual Other. In other words, we assume that the nature of co-present experiences is embodied and enactive, and assume consequently that the Author involved in designing experiences for others in narrative environments, VR or otherwise, can detect and take advantage of similarities between their own subjective-phenomenal simulation and those of the target Experiencer. For the sake of focus, we assign the notion of *empathy* a central position in describing the assumed dynamics of enactive simulation and its psychophysiological, affective-cognitive determinants. Further, our case study, *Booth,* facilitates discussing these theoretical ideas in relation to a simulated co-presence situation within an asylum seeker interview context involving the socioemotional dynamics that we assume to constitute the crucial prerequisite of designing enactive co-presence with an artificial Other. Finally, we envision further studies that involve human co-presence with CGOs in enactive narrative systems in virtual reality installations. Altogether, these efforts aim to create a framework of enactively intelligent artificial character behavior in varying narrative contexts, whether conceived of as an end in itself or as models of enactive human co-presence.

1.1 Theory of Enactive Mind

The general conceptual framework elaborated in this paper assumes that the mind is embodied seamlessly within a situated brain-body system, an idea inherited from the early phenomenological inquiries of Merleau-Ponty (1945/1962). Within this discussion, we trust particularly in the theoretical conceptualization of Varela et al. (1991) from whom we adopt the concept of *enactive mind*, referring to the idea that the said brain–body *system* is further embedded and active within the sociocultural world and therefore inseparable from it. According to this view, the human mind *enacts the world* by continuously reconstructing its situatedness as codeterminate with it (see Ward et al., 2017, for an overview on enactivism).

In order to characterize the relation of an individual mind with other minds of the sociocultural world, Tikka defines the mind's function as *enactive simulation* of others (Tikka, 2005, 2006, 2008, 2022), that is, as the affective-cognitive process that aims at understanding and anticipating behaviors of other beings. Understood in holistic terms, enactive simulation can be assumed to involve a range of simultaneous psychophysiological brain–body dynamics that engage both unconscious and conscious, or affective and cognitive processes, all in inseparable interplay with contextual dynamics of the sociocultural world. We assume that it is enactive simulation of social situations that drives experience not only with human Others, but even when engaged with CGOs in immersive narrative systems. In this sense, the Other is understood as the general signifier for Other beings, whether human or artificial. Due to the lack of direct access to the experience of Others, the idea of simulating Others suggests that an individual mind can project its own embodied experiences and its own enactive functions upon the experience of others.

According to our view, the mind *always enacts within context.* Accordingly, understanding the Other by means of simulation is crucially dependent on contextual frames. As we suggest, what constitutes the determinant context for an Experiencer is a lifetime of embodied socioemotional encounters embedded in socially shareable narratives. We interpret such determinant context as a *temporally extended narrative,* which can be mediated by means of cinema, virtual reality experiences, games, literature, and so on. As such, attempts at modeling such a complex dynamicity that involves biological, physical, and phenomenal domains calls for multidisciplinary descriptive and functional methods. In theorizing about such systems, we rely on a significant body of research in psychology and cognitive neuroscience with focus on mediated narratives in the context of shared experience.

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

1.2 Enactive Narrative Systems and the Artificial Other

With enactive narrative systems we refer to designed technology-implemented setups, often involving Virtual Reality (VR) environments with several subsystems, including computer-generated entities co-constituting the situatedness with human Experiencers. In such systems, the Other that the Experiencer encounters may be modeled as an algorithmic entity with context-dependent behavior where machine learning is implemented. Such an artificial Other would be considered "enactive" if it were to be programmed so that it could model and predict the behavior of other agents (human and virtual), and thereby set expectations about the likelihood of certain events (Tikka et al., 2012; Kauttonen et al., 2014). Behavioral modifications of the Other could be driven by a set of predesigned decisions based on the real-time changes in the psychophysiological measures from the Experiencer, or they might be based on supervised (hybrid intelligence) or unsupervised learning (artificial intelligence). The setting described is in line with "the design of 'Other' Intelligences (OI) . . . coupled with a given observer" (de Melo, 2022, p. 1).

Psychophysiological manifestations of the human experience (e.g., eye movements, heart rate, brain signals) may be analyzed in real time and used to modify the behavior of the virtual Other, which, in turn may then affect the Experiencer in a context-dependent manner, ad infinitum. The collection of larger context-locked psychophysiological time-series data sets could improve the possibility of building a predictive model that would allow the changes in implicit biosignals of the Experiencer to drive an enactive system, including the Other and the narrative context.

This can also be considered as a form of advanced, technologically mediated storytelling, such that it allows for the modification of the narrative context based on real-time psychophysiological changes tracked from the Experiencer (Tikka et al., 2006, 2012; Tikka, 2008; Tikka & Kaipainen, 2014). Alternatively, the data could be modeled upon previously recorded or historical context-dependent data from other human interactions. The aim of such modeling is to gain systemic understanding of context-determined co-presence between a human participant (the Experiencer) and an artificial humanlike person (the Other) in a narrative context.

1.3 Triadic Model of Enactive Co-Presence in Narrative Context

In order to conceptually frame the components of what we consider as narrative system dynamics, we propose a triadic model that describes enactive co-presence in a tripartite manner, as consisting of (1) the human Experiencer, whose behavior is to some extent unique and unpredictable, (2) the Other, another being that is physically or virtually present with the Experiencer, and (3) the narrative Context as the condition of determining the experientially meaningful co-presence of the two in terms of a weighted sequence of prior experiences. It is, as we put forward, the reciprocal systemic co-determination of these three elements that constitutes co-presence (Experiencer–Other, Experiencer–Context, Other–Context).

The experience of co-presence with a virtual Other may involve active cooperation, minimally shared attention towards something, or the opposite—active withdrawal or rejection of others who are physically or virtually present. As we suggest, what constitutes the context is the experiential history of comparable socioemotional encounters, embodied and realized through memory. It should be emphasized that such histories influence not unlike stories, that is, narratives in audiovisual storytelling or literature. Further, sharing a narrative space with virtual Others is not just a mechanistic interaction, but a highly context-dependent enactment comparable to physical "sense of being with another" (Biocca et al., 2003, p. 456); it "renders persons uniquely accessible, available, and subject to one another" (Goffman, 1959, p. 22; see also Oh et al., 2018). In accordance with the view of enactive mind, co-presence is co-constituted in a reciprocal manner by the enacting parties, involving enactive sociocognitive processes (Kahl & Kopp, 2023), to the extent of "participatory sense-making" (De Jaegher & Di Paolo, 2007).

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

De Jaegher and Di Paolo's (2007, pp. 503–504) discussion of *participatory sense-making*, or enactive social emergence, relies on a subtle distinction between human–human encounters versus human–object encounters, where CGOs might be thought to fall, if perhaps hybridly, into the "object" category. The authors note that human sense-making, when it comes to objects, is "complete" or "fully and lawfully constituted" by the Experiencer's perspective, as the object is deemed to lack autonomy (De Jaegher & Di Paolo, 2007, p. 503). On the other hand, an encounter with and involving human Others "demands frequent readjustments of my individual sense-making" based on the inherent autonomy of human beings, "available to me in this way only partially" (De Jaegher & Di Paolo, 2007, pp. 503–504). The authors note,

> We don't experience the other-in-interaction as totally obscure and inaccessible, nor as fully transparent (like an object fully constituted by my sense-making activity), but as something else: a protean pattern with knowable and unknowable surfaces and angles of familiarity that shapeshift as the interaction unfolds (De Jaegher & Di Paolo, 2007, p. 504).

When virtual narratives involving CGOs are encountered within an enactive theoretical framework, the distinction between object and human, transparently accessible and obscure, knowable and unknowable, becomes blurred. Such virtual "objects" tend to redefine causal capacities of what might be thought of as merely *things*, as well as redefine whether what is to be considered an "essential requirement" of "the social" must always require strict autonomy (cf. Jaegher & Di Paolo, 2007, p. 492)—whether human design elements and authorship or virtual Others might inculcate phenomenal traces or echoes of human autonomy capable of modulating identity remains a viable hypothesis.

What is unique about adaptive CGOs within narrative settings is that they appear to operate on a spectrum, depending on narrative context, of having the capacity to evoke a certain "recursive effect" (Jaegher & Di Paolo, 2007, p. 504) or "co-determination of self and other" (Jaegher & Di Paolo, 2007, citing Thompson, 2001; Thompson & Varela, 2001), as a kind of enactive feedback loop (Tikka et al., 2006; Tikka & Kaipainen, 2014), that was generally thought not to be available within the realm of objects. In this way, CGOs are unique from an enactive point of view, in that they appear to have the capacity to simulate and evoke feelings akin to human-to-human interaction, where assumedly, the more capable of interaction—whether through scripts or machine learning adaptation—the more they may engage in "modulation" (De Jaegher & Di Paolo, 2007, p. 504) of the Experiencer's inner world and subvert expectations.

Compared to physical social encounters, the Experiencer may safely step outside of the quotidian realm of socially constructed norms into a constructed world with "more flexible interpretive resources to infer the feelings and thoughts of characters" (Kidd & Castano, 2013, p. 378). In designed virtual experiences, co-presence with the virtual Other may feel less risky—without "potentially threatening consequences" (Kidd & Castano, 2013). That is, the Experiencer can easily step away from engagement with the virtual Other without any consequences or risks of ever encountering this Other again in real life or the future.

1.4 Author's Second-Order Simulation of Co-Presence in Enactive Narrative Systems

When evaluating the enactive model involving virtual Others, it is useful to focus on the design or authorship of the experience, rather than the Experiencer's perspective alone. The concept of second-order authorship, first introduced in the *Enactive Cinema: Simulatorium Eisensteinense* (Tikka, 2008), as well as the more recently elaborated idea of second-order simulation (Tikka, 2022) both rely on the dynamical systems approach to describe the cognitive processes related to designing enactive narrative systems. In our discussion, the term *Author* (singular) is made to reference the entire agency involved with the design of narrative systems, that is, the team including scriptwriters, programmers, sound designers, and animators. What raises the Author's simulation to second-order when designing enactive narrative systems is that their own affective-cognitive experience related to the designed situations and events is applied

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

to simulate the potential Experiencer's (viewer, listener etc.) enactment when encountering artificial Others in varying contexts. In this way, human autonomy is also embedded in the virtual object, allowing for the virtual Other to embody a form of human "alterity," beyond its merely physical behavioral or narrative presence (cf. Zahavi, 2005, as cited in De Jaegher & Di Paolo, 2007, p. 504).

Importantly, second-order simulation assumes, to some extent, symmetrical psychophysiological dynamics between the Author and the Experiencer, as they both separately simulate the personality, behavior, and intentions of the Other inhabiting the narrative space through their embodied enactment. Relying on this assumption, we suggest that the Author executes designer's power over the enactive narrative system defined in terms of the triadic dynamics of co-presence (described above).

The Author's second-order simulation model assumes indirect control over the first-order co-presence between the Experiencer (to some extent predictable) and the Other (to some extent unpredictable). Raised to second order, the simulation model allows the Author to distinguish three manifestations of enactive co-presence between the Experiencer and the Other in the narrative environment: (1) as it is simulated during the authoring process, (2) as enabled by the behavior of the artificial Other, and (3) as enacted in the behavior of the Experiencer.

In our elaboration, the Author's second-order design process relies on enactive simulation of the co-presence between the Experiencer and the Other within the narrative context. Considering an artificial Other, the design process involves enactive simulation of its observable humanlike behaviors across a range of contexts.

2 Related Work

A range of affective-cognitive processes is assumed to enable (1) the Author's second-order simulation of the triadic dynamics of co-presence during the design process, and (2) the Experiencers' enactive simulation of the Other in a shared context in real time during the designed co-presence. Commonly shared knowledge about human affective-cognitive processes is available to the Author, together with the Author's own reflections related to their individual life experiences. However, while conscious simulation of the experiences and behaviors of other entities dominates these experiential processes, the Author needs to consider a range of intrinsic mental processes that may sidetrack the anticipated experience and expected end result.

The theoretical enactive mind framework offers a conceptual foundation for understanding what genuine human behavior and experience of co-presence may entail, which further anticipates and sets the criteria of human-likeness in the artificial Other; however, complexity of the integrated neural, physiological, and phenomenal dynamics pose difficult—maybe even impossible—challenges to exhaustive observation and description using scientific methods. In order to reduce the holistic complexity towards a more observable domain, we choose to consider *empathy* as a key determinant of enactive simulation, as one among many other potential determinants from the rich field of social psychology (e.g., trust, concern of risk, fear). Empathy is not only one of the most relevant aspects of co-presence, but it has also been discussed rather broadly across scientific literature in terms of encounters between humans (see de Waal, 2021; Vollberg & Cikara, 2018) and humans and artificial agents (e.g., robots, Malinowska, 2021).

2.1 Empathy

Empathy can be described as a multidimensional social construct reflecting "reactions of one individual to the observed experiences of another" (Davis, 1983, pp. 113, 123). It engages mental processes of understanding, feeling, sharing, and self–other differentiation (see Håkansson Eklund & Summer Meranius, 2021). Depending on the nuances of an experience, distinct neural networks of empathy (Shamay-Tsoory et al., 2009) or perspective-taking (Healey & Grossman, 2018) may be activated. As shown in a neuroimaging study, where various types of tragic depictions in film sequences were used as stimuli, two differing experiential versions of

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

empathy were mapped to different psychophysiological states: affective empathy related to automated mirroring of another's momentary unfortunate situatedness (e.g., the person is dying now), and cognitive empathy related to mentalizing unfortunate yet distanced situatedness of the other (e.g., the person will die in the future) (Raz & Hendler, 2014). Such affective/cognitive dynamics can be assumed to overlap more or less with other descriptions of psychophysiological dynamics of understanding other people, such as embodied simulation (Gallese, 2005); Theory of Mind (TOM; Frith & Frith, 2005); as well as affective/cognitive perspective-taking (Healey & Grossman, 2018). Undoubtedly, many types of affective-cognitive processes play a part when experiencing co-presence with another person.

The above-mentioned empathy-related processes may be developmentally even preceded by the embodied shared attention mechanism (SAM) (Baron-Cohen & Swettenham, 1996, p. 159), not very different from the joint referential attention concept of *deixis* (Tikka & Kaipainen, 2019). Engaging with new types of contextual situations depicted in designed narratives may involve the Experiencer's body–brain–world system in multiple alternatives, and even conflicting manners, depending on the context (MacDorman, 2019). Several psychological studies confirm that narratives may help to cultivate understanding of other people through perspective-taking (Kidd & Castano, 2013; Mar & Oatley, 2008).

Although not sharing the same sociocultural background, humans nevertheless have relatively similar psychophysiological simulation capabilities warranted by evolution (Chasiotis, 2011; Neumann & Westbury, 2011), which might to some extent allow for transcending individual differences between human agencies that share the same fictional, simulated, or physical context (Tikka, 2022). The contextually induced and, if possible, also temporarily locked dependencies of interpersonal psychological states allow correlating subjective reports and implicit bodily signals—measured by means of neuroimaging, psychophysiology, and affective computing (see Cowley et al., 2016; Jääskeläinen et al., 2020)—to provide access to understanding behavior in relatively complex socioemotional situations.

2.2 In-Group/Out-Group Studies of Empathy

When interpreting empathy in terms of virtual Others in designed contexts, intrinsic intergroup biases may have their say in how the experience of co-presence unfolds. The studies on intergroup biases address perceived context-dependent differences in ethnic, social, behavioral, or other apparent aspects (see Saarinen et al., 2021). Studies have shown that intrinsic bias dynamics may cause lesser degrees of empathy evocation in Experiencers who are perceiving people that differ from them (referred to as members of the out-group), while increased feelings of empathy have been associated with Experiencers' identification with members of what they consider to be an in-group (Saarinen et al., 2021; Bertrand et al., 2018; Zillmann, 1991, pp. 149, 191; Tajfel & Turner, 1986; Jose & Brewer, 1984). This is in line with studies demonstrating that affective empathic responses to perceived pain of another person correlated positively with perceived racial similarities, and vice versa (see Han, 2018).

Automated, unconscious biases in societal decision-making may manifest in several detrimental forms, particularly affecting those most vulnerable (Phillips-Wren et al., 2019; Rousseau & Foxen, 2010). Studies focusing on fairness in asylum-seeking processes reveal that the socially constructed perception of applicants may be reduced to simplified dichotomies between "good and bad, legitimate or illegitimate" (Rousseau & Foxen, 2010, p. 72). This is manifested, for instance, when an immigration officer bases a decision on whether to accept or reject an asylum application based on a stereotype about the credibility of those from certain countries, a phenomenon that has been termed the "culture of disbelief" (Anderson et al., 2014, p. 14). Such "motivated reasoning" is seen to draw support and reinforcement from shared predispositions and biases (Croston & Pederson, 2013; see McNamara & Tikka, 2023).

Having the capacity to adopt differing perspectives, such as attempting to adopt another's positionality, may elicit more (cognitive) empathy, a process relying on reflexivity towards socioculturally learned expectations and behaviors. While cognitive empathy may take on a

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

compensatory function against more automated (and plausibly more negative) affective empathic responses, and thereby instantiate a more egalitarian expression of empathy (Di Palma et al., 2019, p. 2), during conditions of mental pressure or physical duress, people may tend to revert to more biased thought processes rather than perspective-taking (Phillips-Wren et al., 2019, p. 63). As will be discussed in relation to the *Booth* experiment, in-group/out-group biases may function differently, and take on a new context-dependency, when dealing with CGOs or virtual Others.

2.3 Empathy for an Artificial Other in Context

Is it possible that human encounters with artificial Others in shared narrative spaces would follow similar affective-cognitive dynamics of human–human co-presence, such as mirroring, mentalizing, feelings of empathy, and joint attention? In the CGO literature, the phenomenon of Uncanny Valley (Mori et al., 1970/2012) is regularly used to refer to the unbridgeable experiential abyss between interacting with humans or their artificial replicas. However, the evidence from a neuroimaging study by de Borst and de Gelder indicates—in contradiction to the uncanny valley hypothesis—that "the most human-like characters are processed most similarly to human stimuli on a behavioral and neuronal level" (2015, p. 9). But the social roboticists Jesse Fox and Andrew Gambino warn that theories of human–human relationships might not be automatically suitable for examining relationships between humans and artificial Others. What differences rule and whether they offer advantages in some sense remains a fertile and timely ground for study. de Borst and de Gelder note that "an expansion of neuroimaging studies that compare different properties of avatar and human faces and investigate its underlying brain activity in an experimentally and statistically well constructed way is needed in order to further understand its underlying neuronal mechanisms" as there are a dearth of studies on the subject (2015, p. 9).

From the Experiencer's point of view, the sense of co-presence with virtual Others involves perceived mutual awareness where the Experiencer should feel reciprocality in the behavior of the virtual Other (Pianzola et al., 2021). However, the CGOs have been suggested not to be capable of activating parts of the brain associated with cognitive empathy similarly as interaction with actual humans (Han et al., 2005). Against the findings of Han et al., there are neuroimaging studies that have suggested quite the opposite, that "expressions of emotions in human-like avatars can be perceived similarly to human emotions, with corresponding behavioral, physiological and neuronal activations, with exception of physical dissimilarities'' (de Borst & de Gelder, 2015, p. 1). What should one make of the apparently conflicting results?

There is reason to assume that the sense of co-presence is likely multidimensional, allowing an as-if sense of being co-present with a CGO, while simultaneously recognizing the artificiality on some other experiential dimension. Perhaps it is enough that the Experiencer's construction of reciprocality is supported, as is commonly the case with human–computer interfacing, where even a smiley may suffice to communicate enough reciprocality for the purpose. Perceiving minimal social responsiveness of a virtual Other has been found to elicit the experience of co-presence, even without direct interaction with it (Garau et al., 2005, p. 104). Socially relevant "specific gestures, the timing of mutual gaze, the coordination of verbal and nonverbal activities, etc." of the artificial Other have been shown to elicit a higher level of experienced psychological involvement from the Experiencer (Bente et al., 2004, p. 55).

An important suggestion is that while humans are subject to a range of intrinsic "noise," such as biases and stereotypes that influence human–human co-presence, it may be that social robots (that resemble humans) might not evoke such sociocultural noise by design (Fox & Gambino, 2021, p. 298). This might imply that in human–robot interaction in which robots are mimicking humans, but are still perceived as functional machines, the artificially induced as-if situatedness could override the sociocultural noise immanent in human–human co-presence. To avoid the folk psychological pitfall of assuming that artificial agents (robots) could be perceived as actual humans, Seibt proposed applying five modes

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

of "as if" simulations, namely, "functionally replicating, imitating, mimicking, displaying, and approximating" in order to describe the sociality in human–robot interactions (Seibt, 2017, pp. 21–27). Such simulation of the "as-if" condition may involve deliberate anthropomorphizing (Damiano & Dumouchel, 2018), that is, playing with the idea that an artificial entity (in narrative context) shares human-like traits such as hidden motivations, intentions, or perhaps even a life story to explain why the Other is co-present in the situation (Tikka et al., 2022).

In a study conducted in a clinical interview setting, tested patients demonstrated "less hesitancy to disclose information and showed their emotions more openly" when they understood that the virtual interviewer was fully automated vs. controlled by a human (Lucas et al., 2014, p. 94). Such particular cases show that context may largely define how the virtual Other is perceived. One might assume that virtual Others would fall squarely within a so-called out-group, and thereby be subject to corresponding intergroup biases, as has been shown to be the case in human–human interaction. However, due to "artificial" characteristics of a virtual Other in comparison to, for instance, human interviewers, an Experiencer may assume that biases that might be at play in human–human intergroup judgments and perspectives are not present in encounters with machines.

In the next section, we will discuss the case study *Booth*, an experiment that was designed to study Experiencers' affective-cognitive processes while engaging with virtual asylum seekers and their narratives. This setting exemplifies the triadic co-presence which is mutually co-constituted by the Experiencer, the virtual Other, and the Context. The empathetic responses to the virtual Other are understood as markers for the effectiveness of such settings to elicit empathic concerns, which influence cognitive perspectives taken by Experiencers.

Based on the discussion above, we expected that the Experiencers' empathetic response to the artificial Other would be manyfold, that is, in some aspects similar to that of human Others, some not. Our hypothesis (H) was that based on the capabilities of enactive simulation, the Author would be able to manipulate the judgments of the Experiencer qua immigration officers by priming them with different contextual narratives.

3 Methodology: The Booth Case Experiment

The *Booth* experiment was set up in order to study the Experiencers' empathic responses to two Others, virtual male Asylum Seekers with different narratives that contextualized their justification for asylum. In the experiment, the situatedness of both the Asylum Seeker and the volunteering Experiencer was created by introducing a role-play scenario where Experiencers were asked to assume the role of an immigration officer who has to decide whether to accept or reject the asylum request.

The virtual Asylum Seekers represented different ethnicities, one Middle Eastern and the other North African, as well as two different professions, one a journalist who has been forced to leave his country because of his political views, and the other a medical doctor, who chose to leave his home voluntarily. For the experiment, voice-overs were recorded that describe in chronological order the events that led each to leave their home country and apply for asylum abroad. The narratives were built around the most common questions immigration officers ask refugees, such as how they traveled to the checkpoint and what their reasons were for leaving their home countries.

The events in both narratives were synchronized with one another based on emotional content (e.g., saying goodbye to family; witnessing a horrifying event during the refugee journey) and facts (e.g., professional position; the route of the refugee journey) of the unfolding story. One dynamically unfolding facial behavior was pre-recorded using iPhone's facial motion capture and applied to both Asylum Seekers using Unreal Engine. The same facial behavior was designed to be in synchrony with both prerecorded narratives in order to limit study variables to the two ethnicities of the Asylum Seekers, event variations in their narratives, and the two videos used as (negative/positive) priming conditions.

In the experimental setting, the Asylum Seekers' dynamically animated face was shown on the computer

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

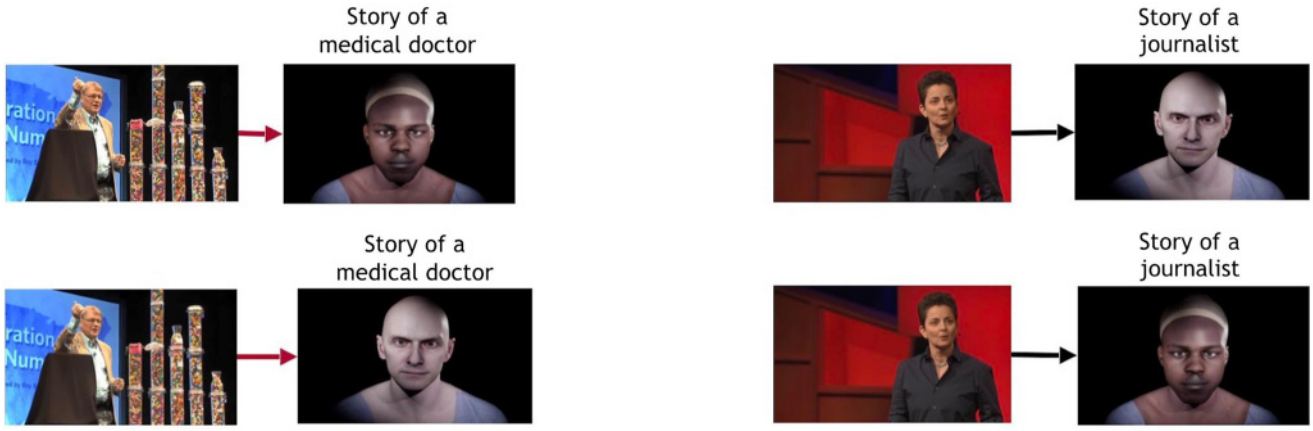


Figure 1. *The two priming videos were selected with two opposing functions in mind: (1) to elicit cognitive empathy but to direct towards rejection decision, and (2) to elicit affective empathy and direct towards acceptance (videos with blue and red background, respectively). (Left video adapted from "Immigration World Poverty, Immigration & Gumballs." Presented by author/journalist Roy Beck. https://youtu.be/KCcFNL7EmwY. Right video adapted from "Don't feel sorry for refugees – believe in them." TED talk presented by Luma Mufleh. https://youtu.be/6wNif5SIN08.)*

screen while his voice-over was played. After each encounter with the Asylum Seeker, the Experiencer freely determined their acceptance or rejection. The decision process of refugee status acceptance or rejection was not based on any specific legal standard; instead, it relied on the Experiencers' personal inclination and empathy-based decision-making abilities.

3.1 Experiment Setup

3.1.1 Experiencers. The *Booth* involved 13 men and 9 women who were on average 25.8 years old. The sample group consisted of Estonian nationals who were either students or faculty at Tallinn University. *As-if-role:* The Experiencers were instructed to adopt the role of an immigration officer. *Pre-experimental IRI-questionnaire:* Prior to the experiment, each participant filled out a questionnaire designed to measure their level for empathy predisposition, using the Interpersonal Reactivity Index (IRI) (Davis, 1980). Based on pre-experiment IRI results, the Experiencers were divided into two groups that were primed with either negative or positive videos on accepting asylum seekers into a host country.

3.1.2 Two Priming Videos. The videos were selected with two opposing functions in mind: (1) to elicit cognitive empathy *but* to direct towards rejection decision, and (2) to elicit affective empathy *and* direct towards acceptance (see Figure 1, videos with blue and red background, respectively). In the critical video on immigration the speaker argued, using visual quantifiables, that help benefiting only a few individuals does not remove the root cause of immigration in their home country—poverty. This priming video was always followed by the asylum story of a medical doctor. In the other talk, the speaker's autobiographical message was emotionally loaded and encouraged people to help asylum seekers even if it meant making a difference for just one person. This priming video was always followed by the asylum story of a journalist (see Figure 1).

3.1.3 Experiment Protocol. Each session began with the Experiencer watching a priming talk which expressed one of the two views on immigration. Following the priming talk, the Experiencer faced a virtual asylum seeker on a computer screen while listening to their story. After the encounter, the Experiencer (role-playing as an immigration officer) was asked to either grant or deny asylum (see Figure 2). After a short break, the Experiencer viewed a different priming talk presenting another point of view on immigration. Similar to the first part of the experimental session, another virtual Asylum Seeker with their narrative was presented and the

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

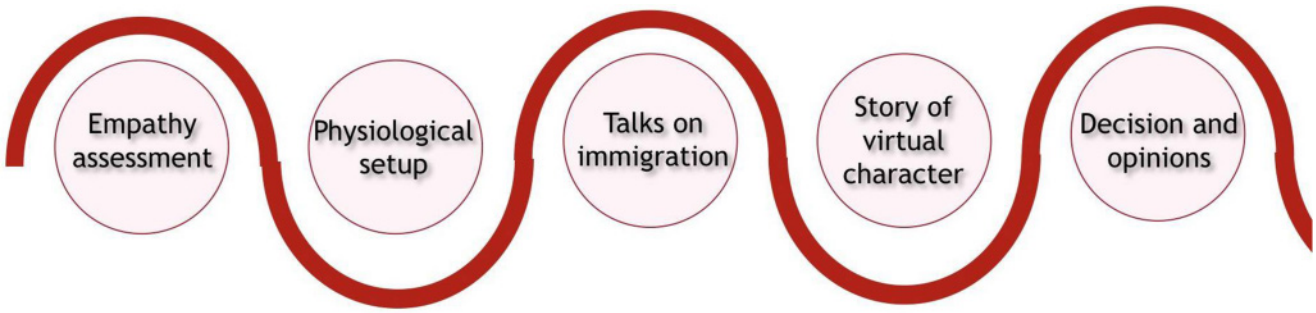


Figure 2. *The graph shows the timeline for the study protocol. After the IRI empathy assessment, a selection of biosensors was attached to the Experiencer's body. The session started with a priming video on the screen, followed by the video that depicts the Asylum Seeker's animated face and his narrative as a voice-over. After each session, the Experiencer was asked to decide whether to accept or reject the asylum request.*

Experiencer was asked to either grant or deny asylum. All Experiencers saw both priming videos followed by the refugee narrative once. The ethnic CGOs were counterbalanced, as well as which priming video was shown first. *Experiment duration*: Each session lasted approximately 2 hours.

3.2 Empathic Concern and Perspective Taking

In order to simplify the experiential complexity related to the topic, a selection of affective-cognitive functions of the socioculturally situated human mind were chosen as determinants of co-presence: empathy, perspective-taking, intrinsic biases, and their context-dependency. Prior to the experiment, each participant filled out a questionnaire designed to measure their level for empathy predisposition, using the Interpersonal Reactivity Index (IRI) (Davis, 1980). The IRI employs a Likert scale for participants to rate questions; answers range from "does not describe me very well" to "describes me very well." The 28 statements in the questionnaire are associated with four different aspects of empathy (perspective-taking, fantasy, empathic concern, and personal distress), with each aspect being assigned seven statements, respectively. The two categories of interest in the *Booth* were empathic concern (EC) and perspective-taking (PT). EC is viewed as emotional or affective empathy as it "inquires about respondents' feelings of warmth, compassion, and concern for others" (Davis, 1983, p. 113). The PT scale is meant to evaluate respondents' ability to "adopt the perspectives of other people and see things from their point of view" (Davis, 1980, p. 2).

The statements that measure EC ability were as follows (Davis, 1980, p. 7):

1. I often have tender, concerned feelings for people less fortunate than me.
2. Sometimes I don't feel very sorry for other people when they are having problems.
3. When I see someone being taken advantage of, I feel kind of protective towards them.
4. Other people's misfortunes do not usually disturb me a great deal.
5. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
6. I am often quite touched by things that I see happen.
7. I would describe myself as a pretty soft-hearted person.

The statements that measure PT ability were as follows (Davis, 1980, 7):

1. I sometimes find it difficult to see things from the "other guy's" point of view.
2. I try to look at everybody's side of a disagreement before I make a decision.
3. I sometimes try to understand my friends better by imagining how things look from their perspective.

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

4. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
5. I believe that there are two sides to every question and try to look at them both.
6. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
7. Before criticizing somebody, I try to imagine how I would feel if I were in their place.

4 Results and Discussion

During the *Booth* experiment, Experiencers' decision-making reflected a high degree of empathy, with the vast majority choosing to grant asylum, regardless of priming video content, Asylum Seeker's ethnicity, or narrative content. Priming the Experiencer with a negative video prior to engaging with the Other did not reduce reported empathy with the asylum narrative. While the Experiencers of the study had more or less average predispositions for empathy, they voted overwhelmingly in the positive.

The analysis of the IRI questionnaire responses showed that the mean empathic concern (EC) and perspective-taking (PT) scores for the sample group were both 3.73. When comparing those scores to the numbers generated as a result of large population studies (O'Brien et al., 2013, as cited in Konrath, 2013; Sommerlad et al., 2021), it is evident that participants in the *Booth* experiment scored similarly, which implies that the sample of the study was generally not more or less predisposed towards empathy than what seems to be the average. For instance, in an online survey of the general American adult population (GSS), which involved 81,754 people, the PT mean score for the group was 3.66, while the mean EC group score was 3.77 (O'Brien et al., 2013, as cited in Konrath, 2013). In a UK study, which involved 30,033 participants and was meant to be a representative study of the whole population, the mean empathic concern score for the group was 3.86, and perspective-taking was 3.57 (Sommerlad et al., 2021). The only noticeable differences from earlier studies were that the PT predisposition score for the *Booth* experiment participants was slightly higher than participants in the other large-scale population studies, while the EC score was slightly lower. Interestingly, as a result of confirmatory factor analysis, Wang et al. (2020) inferred that the PT scale should be used as a standalone indicator of cognitive empathy, the approach relevant for our discussion.

In this section, we will reflect retrospectively on enactive simulation of second-order authoring of the *Booth* experiment in terms of affective-cognitive and sociocultural biases, be these explicit by design, intrinsic, or by assumption. The future applications of enactive concepts and the proposed triadic model of enactive co-presence are further elaborated.

4.1 Second-Order Authoring of Co-Presence

The *Booth* experiment was set up in order to simulate situations where the Experiencer and the asylum seeker encounter each other and are seated face-to-face in an immigration interview room, and where both are present while a previously recorded narrative of the asylum seeker plays. According to our assumption, the three components determining the nature of the co-presence were (1) the Experiencer taking the role of immigration officer, (2) the artificial Asylum Seeker being virtually present with the Experiencer, and (3) the refugee journey narration as the conditioning context. The Experiencer was given the role of immigration officer responsible for making a fair decision about whether to accept or reject the asylum claim.

Our focus in this paper has been on the second-order simulation, that is, the designer's means to understand the Other by simulating the Experiencer's affective-cognitive responses to a CGO (in this case the immigration officer). With the purpose of probing how the experience of co-presence could be influenced, we primed the *Booth* experiment by first exposing the Experiencer to one of the two alternative videos which were thought to have a tendency to evoke potential intrinsic biases, such as that the narrative was purposefully tailored, or perhaps that the narrative would contain untrue elements. The narratives left open the possibility for

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

elicitation of racial, cultural, gender-based, or politically oriented biases that were thought to have the capacity to potentially influence the officer's decision. The Experiencer's eventual decision in the role of an immigration officer was expected to reflect their own accumulated experience by means of enactive simulation, integrating different viewpoints involved in the asylum-seeking situation, such as simulating the embodiedness and situatedness of the artificial Other to whom their own past Experiences are projected.

In the *Booth*, the Asylum Seekers were modeled to represent different ethnicities than the Experiencers, who were of Caucasian ethnicity. While perceived outgroup membership has been associated in previous studies with a tendency to negatively influence empathic decision-making (Saarinen et al., 2021), this was not confirmed by our case.

Against our hypothesis, the majority (37 out of 44) of the Experiencers' final decisions were to accept the asylum request based on listening to the Asylum Seeker's prerecorded narrative while watching their animated face on the screen, and regardless of viewing either "negative" *or* "positive" priming videos with narrative justification for seeking asylum (whether seemingly "legitimate" or "illegitimate"). One possible reason for the uniformity may relate to the holistic interplay between affective and cognitive simulation processes, considering the two as-if conditions determining the study: the Experiencer was asked (1) to assume the position of an immigration officer (role-play condition), and (2) to consider the CGO as representing an actual asylum seeker (as-if-real-human condition).

One possible explanation is that awareness of these artificial as-if conditions caused Experiencers to adopt a more conscious strategy of cognitive empathy, resulting in a higher disposition of acceptance (cognitive empathy) in this study than initially expected by the designers of the study. Cognitive empathy (or Theory of Mind–based empathy, TOM) has been shown to be evoked or appealed to when an Experiencer is able to encounter a scene from a more distanced third-person stance (see Raz & Hendler, 2014). It may thus be the case that the experience of co-presence with a virtual Asylum Seeker in narrative context may in some instances create a greater tendency to evoke distancing or "defamiliarization" (see Timmermans & Tavory, 2012, p. 180), and thus elicit more cognitive-based empathy. According to a number of neuroimaging studies, the functional neural networks that are related to cognitive empathy, as well as related social perception and mentalizing, may not be as strongly activated by understanding a depicted character's "intentional stance" from movement as compared to live-action characters (Mar et al., 2007, p. 199; Han et al., 2005; Damiano & Dumouchel, 2018). While these studies did not discuss TOM-based empathy in particular, it is tempting to speculate whether this cognitively motivated mode of empathy depicted by animated characters might also mitigate more general findings.

Encounters with artificial Others might invite comparison to the act of mentalizing that has been described in reading fictional stories about others: Engaging with narratives puts the Experiencer into a space where she is more apt to fall naturally into the activities of "mind-reading and character construction" (Kidd & Castano, 2013, p. 377). Were this the case, it would follow that such encounters and contexts with such fictional spaces or artificial Others would provide affordances for increasing moral understanding by eliciting an environment for perspective-taking. This would support the conclusion that a learned conscious mentalizing form of empathy might work to "compensate" for or neutralize perceived differences encountered in the real world (see Di Palma et al., 2019).

It may be that the artificial Other evokes distancing of the type where the artificiality of the Other defamiliarizes the Experiencer from the concept of difference and identification itself, and in this way cognitive perspective-taking may help the Experiencer to reinterpret biases from a simulated third-person view. That the virtual asylum seekers in the *Booth* experiment are not perceived as actual humans, but rather as their mediated representations, allows for co-presence to be experienced as less risky "without facing the potentially threatening consequences of that engagement," as suggested by Kidd and Castano (2013, p. 378).

It could be that the narrative told by the voice of the virtual asylum seeker, in conjunction with the face-to-face encounter with the virtual Other, may allow the

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

Experiencer to safely step outside of the quotidian realm of socially constructed norms wherein "experiences may be scripted by convention and informed by stereotypes" into a world that may provide "more flexible interpretive resources to infer the feelings and thoughts of characters" (Kidd & Castano, 2013, p. 378). Such an encounter creates a meta space for the *enacted idea of asylum seekers* in general and for enacted co-presence with an artificial asylum seeker in particular, yet, without having to face real decision-making or survival consequences.

From the enactive point of view, De Jaegher and Di Paolo (2007, p. 487) have defined an "autonomous system" as a human system "composed of several processes that actively generate and sustain an identity under precarious conditions." While virtual or artificial Others would potentially tend to fall outside of the definition of autonomous as they may be classified as objects, rather than living beings, it is perhaps precisely that they lack autonomy, but still take on human traits from second-order authorship or designership, that they trigger questioning or reevaluation of identity for their Experiencers.

4.2 Limitations of the Work

We acknowledge that the topic remains far from exhausted. We chose to use as a case study the *Booth,* an experiment that we had earlier designed for studying empathy felt towards CGOs, involving computer-generated characters standing for an Other, which in this case represented asylum seekers in a simulated asylum interview. Doing this, we ourselves were in the role of the designer, that is, the second-order author with control of all of the three elements: the Experiencer whose behavior we attempted to simulate, the artificial Other whose behavior we had created, and a contextualizing narrative that we could vary. With regard to the latter, we had the primary hypothesis that we would be able to influence the Experience and consequently the decision of the Experiencer by varying contextually the priming video narrative.

One limitation was that our setting was not representative of a genuinely enactive adaptive dynamics but a simplification, where neither the narrative elements nor the Asylum Seekers' behaviors were adaptive to the modifications in the Experiencer's psychophysiology. A technology-related limitation of the setting was that the virtual Asylum Seekers' narratives were not lip-synced, but instead, both the Experiencer and the Asylum Seeker were as-if co-listening to the refugee narrative, and estimating it from their different perspectives. An epistemological limitation of the work relates to the radical simplification of the asylum context, in which the asylum seekers often do legitimately fear for their safety, which is likely to influence their behavior and influence their strategy in the interview.

We further acknowledge another limitation related to the Experiencers, who were uninitiated participants rather than professional immigration officers. Our participants were university students and staff without adequate information or education on the legal decision-making protocols conducted in actual immigration settings. Their uniform responses may be due to true empathy without representing the immigrant officer's behavior who struggles to judge whether the legal criterion of "well-founded fear" is satisfied or not. At the same time, this selection of informants may be not unlike jury members in trials, likewise driven by other influences besides the legal.

We have addressed both legal and practical issues of implementation of advanced machine-learning methods and adaptive virtual agents in border control processes elsewhere (McNamara & Tikka, 2023). Before concluding, in the next section, we address future potentials of second-order authoring.

4.3 Future Potentials of Second-Order Authoring of Enactive Narrative Systems

In the near future, the ever-increasing computational power of data generation and management will allow for building increasingly immersive adaptive environments inhabited with hyperrealistic artificial characters.

The triadic model of enactive co-presence has been described as constituting reciprocal feedback loops between the Experiencer, the Other, and their

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

co-situatedness in context. Our focus has been on the agency designing such systems, referred to as second-order authorship. We suggest that the related function of enactive second-order simulation is in play throughout design processes for dynamical systems in which narrative contexts unfold in time, and which involve humanlike Others with self-determining behaviors, with the two in co-constitutive relation to one another. The *Booth* experiment exemplified one type of design for studying empathy-driven enactive co-presence in narrative systems. While the results were somewhat unexpected and inconclusive, they may nevertheless guide the design of further experiments that aim to test some of the hypotheses implied, in particular related to biases and affective-cognitive (intertwined) dispositions. In such settings, psychophysiological variables can be used to develop a basic precursor model of empathic response elicited in Experiencers based on the context-dependent behaviors of a virtual Other. Once identified, the neural correlates of affective and cognitive states elicited in specific contexts may theoretically drive empathic elicitation of experiences prompted by dynamical behaviors of any programmable and adaptable virtual Other in similar contexts and contribute to the toolbox of co-presence Designers.

Psychophysiological data can be used to guide machine-learning algorithms. For example, the system could be designed to elicit a maximum amount of empathy towards the virtual Other, and the level of empathy detected from the Experiencer could be fed back as a "reward" to the underlying machine learning model. So, analogous to chess-playing artificial intelligence that learns by receiving rewards for winning strategies, the enactive AI would learn strategies to elicit different desired states in the Experiencers' situatedness. These would be based on real-time analysis of emotional responses of the Experiencer, feeding this data back to the Other, in order for the behavior of the Other to retrigger new responses in the Experiencer, now plausibly generating a new contextual situation.

We consider further elaboration of our vision of enactive narrative systems towards a direction that would allow dynamical unfolding of co-presence between an Experiencer and the virtual Other, in terms of mutually determined dynamical co-constitution of shared situatedness. Such a system would rely on real-time reading of the changes in the Experiencer's psychophysiological behavior (indicating modifications in the state of the human brain-body system). Undoubtedly, a number of ethical issues will arise with these practices. It may be considered our contribution that we have reviewed a number of influences that may be involved in the affect-evoking systems of co-presence. As they cannot be exhausted here, they deserve a broad discussion on their own.

5 Conclusions

We have tentatively discussed what we consider as second-order simulation of enactive co-presence of Experiencer and the virtual Other, and in particular the role of narrative context in modulating it. Our approach builds on the theory of enactive mind as a theoretical starting point of a discussion of various affective-cognitive processes involved in designing immersive narrative systems for human–machine co-presence. As a conceptual model for the discussion, a triadic model of enactive co-presence between (1) virtual characters and (2) humans in particular (3) contexts was put forward. We interpreted and operationalized the experience-determining context as a *temporally extended narrative,* mediated in the present case by means of encounters with virtual on-screen avatars or Others. Here, the concept of *enactive simulation,* relying on accumulated embodied and involved experience, served as a means of understanding an Other, be it another human or an artificial humanlike character. Further, the concept of *second-order simulation* refers to the cognitive-affective means of the designer of such sociotechnological systems, that is, simulating how the Experiencers simulate the Other as the guide.

For the sake of focus, we chose to examine *empathy* out of a number of other affective dispositions, such as hate or trust. A special position was given to empathy in *enactive simulation*, not least thanks to the observable psychophysiological manifestations of this affective-cognitive phenomenon, but also because

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

empathy appears to function as a key requisite of being able to simulate an Other.

Our experimental case study, the *Booth*, allowed us to discuss the proposed triadic model in terms of affective-cognitive and sociocultural system dynamics. We generally expected that the Experiencers' empathetic response to the CGO would be complex and manyfold, that is, in some aspects similar to that to human Others, while in some following other regularities instead. Partially addressing this complexity, we had the hypothesis that we as designers would be able to manipulate the empathy in the experience of co-presence, and guide the consequent asylum judgments of the Experiencer (acting as immigration officers) by priming them with different contextual narratives. However, the results did not directly support this. The majority of Experiencers turned out to support asylum regardless of the priming. This might—in one way or another—be a consequence of the complexity and multilayer nature of the empathy evoked by CGOs in various contexts: Perhaps empathy and more reserved attitudes can overlap? The falsification of the hypotheses should simply be interpreted as a design implication in a series of iterative trial-and-error, as is common in the creative domain. For example, a reference to the legal criterion of "well-foundedness of fear" might suffice to turn the empathy toward a more cognitive direction. Overall, the experiment explored and implied aspects of the complexity and guided further hypothesis formation.

Finally, learning from the experiment, we can envision further studies that involve human co-presence with virtual characters in immersive contexts. We trust that our work contributes to the understanding of immersive narrative systems such as interactive games, interactive media applications, interactive cinema, and interactive VR installations with artistic purposes. The emerging field will have a number of repercussions within and upon society parallel to social robots, currently being introduced even for various societal tasks in addition to entertainment, art, and games. The modeling approach accumulates knowledge about sociopsychological encounters that would otherwise be difficult to observe systematically, or might require ethically questionable interventions. The knowledge will contribute to the education of designers of such systems also by identifying and weighting key elements of the enactive dynamics.

Acknowledgments

This work has been supported by the EU Mobilitas Pluss Top Researcher Grant (MOBTT90), Estonian Research Council.

REFERENCES

Anderson, J., Hollaus, J., Lindsay, A., & Williamson, C. (2014). The culture of disbelief: An ethnographic approach to understanding an under-theorised concept in the UK asylum system. *Refugee Studies Centre*, *102,* 1–20.

Baron-Cohen, S., & Swettenham, J. (1996). The relationship between SAM and ToMM: Two hypotheses. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind*. Cambridge University Press.

Bente, G., Rüggenberg, S., & Krämer, N. C. (2004). Social presence and interpersonal trust in avatar-based, collaborative net-communications. *Proceedings of the 7th Annual International Workshop Presence*, pp. 54–61.

Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., & Zenasni, F. (2018). Learning empathy through virtual reality: Multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI*, *5.* 10.3389/frobt.2018.00026

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments, 12*(5), 456–480. 10.1162/105474603322761270

Chasiotis, A. (2011). Evolution and culture. *Online Readings in Psychology and Culture, 9*(1), 1–17. 10.9707/2307-0919.1101

Cowley, B. U., Filetti, M., Lukander, K., Torniainen, J., Helenius, A., Ahonen, L., . . . & Jacucci, G. (2016). The psychophysiology primer: A guide to methods and a broad review with a focus on human–computer interaction. *Foundations and Trends in Human-Computer Interaction*, *9*(3–4), 151–308. 10.1561/1100000065

Croston, J., & Pedersen, A. (2013). "Tell me what I want to hear": Motivated recall and attributions in media regarding asylum seekers. *Australian Journal of Psychology*, *65*(2), 124–133. 10.1111/ajpy.12012

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human–robot co-evolution. *Frontiers in Psychology, 9*. 10.3389/fpsyg.2018.00468

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85. https://www.uv.es/~friasnav/Davis_1980.pdf

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. 10.1037/0022-3514.44.1.113

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*. 10.3389/fpsyg.2015.00576

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507. 10.1007/s11097-007-9076-9

de Melo, A. T. (2022). Conceptualizing and designing (co)augmented intelligence(s) enacting complex thinking: Morphological constraints, challenges and implications. In *Proceedings of The 2021 Summit of the International Society for the Study of Information*. 10.3390/proceedings2022081085

de Waal, F. B. M. (2021). Empathy, the umbrella term. *Neuroscience and Biobehavioral Reviews*, *129*, 180–181. 10.1016/j.neubiorev.2021.07.034

Di Palma, M., Arcangeli, E., Lattanzi, D., Gabbiadini, A., Galluci, M., Cuppini, R., . . . & Berlinger, M. (2019). Heart Rate Variability reveals the fight between racially biased and politically correct behavior. *Nature Scientific Reports*, *9*(1), 11532. 10.1038/s41598-019-47888-w

Fox, J., & Gambino, A. (2021). Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychology, Behavior, and Social Networking, 24*(5), 294–299. 10.1089/cyber.2020.0181

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology, 15*(17), 644–645. 10.1016/j.cub.2005.08.041

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences, 4*(1), 23–48. 10.1007/s11097-005-4737-z

Garau, M., Slater, M., Pertaub, D.-P., & Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments, 14*(1), 104–116. 10.1162/1054746053890242

Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.

Han, S. (2018). Neurocognitive basis of racial ingroup bias in empathy. *Trends in Cognitive Sciences, 22*(5), 400–421. 10.1016/j.tics.2018.02.013

Han, S., Jiang, Y., Humphreys, G. W., Zhou, T., & Cai, P. (2005). Distinct neural substrates for the perception of real and virtual worlds. *NeuroImage 24*(3), 928–935. 10.1016/j.neuroimage.2004.09.046

Håkansson Eklund, J., & Summer Meranius, M. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling, 104*(2), 300–307. 10.1016/j.pec.2020.08.022

Healey, M. L., & Grossman, M. (2018). Cognitive and affective perspective-taking: Evidence for shared and dissociable anatomical substrates. *Frontiers in Neurology*, *9*. 10.3389/fneur.2018.00491

Jääskeläinen, I. P., Klucharev, V., Panidi, K., & Shestakova, A. N. (2020). Neural processing of narratives: From individual processing to viral propagation. *Frontiers in Human Neuroscience, 14*, 253. 10.3389/fnhum.2020.00253

Jose, P. E., & Brewer, W. F. (1984). Development of story liking: Character identification, suspense, and outcome resolution. *Developmental Psychology*, *20*(5), 911–924. 10.1037/0012-1649.20.5.911

Kahl, S., & Kopp, S. (2023). Intertwining the social and the cognitive loops: Socially enactive cognition for human-compatible interactive systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, *378*(1875), 20210474. 10.1098/rstb.2021.0474

Kauttonen, J., Kaipainen, M., & Tikka, P. (2014). Model of narrative nowness for neurocinematic experiments. *Proceedings of the 5th Workshop on Computational Models of Narrative*, pp. 77–87. 10.4230/OASIcs.CMN.2014.77

Kidd, D., & Castano, E. (2013). Reading literary fiction improves Theory of Mind. *Science*, *342*(6156), 337–380. 10.1126/science.1239918

Konrath, S. (2013). Critical synthesis package: Interpersonal Reactivity Index (IRI). *MedEdPORTAL*. 10.15766/mep_2374-8265.9596

Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. 10.1016/j.chb.2014.04.043

MacDorman, K. (2019). In the uncanny valley, transportation predicts narrative enjoyment more than empathy, but only

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

for the tragic hero. *Computers in Human Behavior, 94*, 140–153. 10.1016/j.chb.2019.01.011

Malinowska, J. K. (2021). What does it mean to empathise with a robot? *Minds and Machines, 31*, 361–376. 10.1007/s11023-021-09558-7

Mar, R. A., Kelley, W. M., Heatherton, T. F., & Macrae, N. (2007). Detecting agency for the biological motion of veridical *vs* animated agents. *Social Cognitive and Affective Neuroscience*, *2*(3), 199–205. 10.1093/scan/nsm011

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*(3), 173–192. 10.1111/j.1745-6924.2008.00073.x

McNamara, R. G., & Tikka, P. (2023). Well-founded fear of algorithms or algorithms of well-founded fear? Hybrid intelligence in automated asylum seeker interviews. *Journal of Refugee Studies,* feac067. 10.1093/jrs/feac067

Merleau-Ponty, M. (1945/1962). *Phenomenology of perception*. Routledge & Kegan Paul.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. 10.1109/MRA.2012.2192811

Neumann, D. L., & Westbury, H. R. (2011). The psychophysiological measurement of empathy. In D. J. Scapaletti (Ed.), *Psychology of empathy*, (pp. 119–142). Nova Science Publishers.

O'Brien, E., Konrath, S. H., Grühn, D., & Hagen, A. L. (2013). Empathic concern and perspective taking: Linear and quadratic effects of age across the adult life span. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 68*(2), 168–175. 10.1093/geronb/gbs055

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, *5*. 10.3389/frobt.2018.00114

Phillips-Wren, G., Power, D. J., & Mora, M. (2019). Cognitive bias, decision styles, and risk attitudes in decision making and DSS. *Journal of Decision Systems*, *28*(2), 63–66. 10.1080/12460125.2019.1646509

Pianzola, F., Riva, G., Kukkonen, K., & Mantovani, F. (2021). Presence, flow, and narrative absorption: an interdisciplinary theoretical exploration with a new spatiotemporal integrated model based on predictive processing. *Open Research Europe*, 1–37. 10.12688/openreseurope.13193.2

Raz, G., & Hendler, T. (2014). Forking cinematic paths to the self: Neurocinematically informed model of empathy in motion pictures. *Projections: The Journal for Movies and Mind, 8*(2), 89–114. 10.3167/proj.2014.080206

Rousseau, C., & Foxen, P. (2010). "Look me in the eye": Empathy and the transmission of trauma in the refugee determination process. *Transcultural Psychiatry*, *47*(1), 70–92. 10.1177/1363461510362338

Saarinen, A., Jääskeläinen, I. P., Harjunen, V., Keltikangas-Järvinen, L., Jasinskaja-Lahti, I., & Ravaja, N. (2021). Neural basis of in-group bias and prejudices: A systematic meta-analysis. *Neuroscience and Biobehavioral Reviews*, *131*, 1214–1227. 10.1016/j.neubiorev.2021.10.027

Seibt, J. (2017). Toward an ontology of simulated social interaction: Varieties of the "as-if" for robots and humans. In R. Hakli & J. Seibt (Eds.), *Sociality and normativity for robots: Philosophical inquiries into human-robot interactions* (pp. 11–39). Springer. 10.1007/978-3-319-53133-5_2

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain: A Journal of Neurology*, *132*(3), 617–627. 10.1093/brain/awn279

Sommerlad, A., Huntlay, J., Livingston, G., Rankin, K. P., & Fancourt, D. (2021). Empathy and its associations with age and sociodemographic characteristics in a large UK population sample. *PLoS ONE*, 16(9). 10.1371/journal.pone.0257557

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *The psychology of intergroup relations* (pp. 7–24). Nelson-Hall.

Thompson, E. (2001). Empathy and consciousness. *Journal of Consciousness Studies, 8,* 1–32.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*, 418–425. 10.1016/S1364-6613(00)01750-2

Tikka, P. (2022). Enactive authorship: Second-order simulation of the viewer experience—A neurocinematic approach. *Projections: The Journal for Movies & Mind*, *16*(1), 47–66. 10.3167/proj.2022.160104.

Tikka, P., & Kaipainen, M. (2019). Intersubjectivity, idiosyncrasy, and narrative deixis: A neurocinematic approach. In M. Grishakova & M. Poulaki (Eds.), *Narrative complexity cognition, embodiment, evolution* (pp. 314–337). University of Nebraska Press.

Tikka, P., & Kaipainen, M. Y. (2014). From naturalistic neuroscience to modeling radical embodiment with narrative

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023

4dec60fe7f0c8664ef0b2d39: D26102300517ZX00

enactive systems. *Frontiers in Human Neuroscience, 8,* 794. 10.3389/fnhum.2014.00794

Tikka, P., Väljamäe, A., de Borst, A. W., Pugliese, R., Ravaja, N., Kaipainen, M., & Takala, T. (2012). Enactive cinema paves way for understanding complex real-time social interaction in neuroimaging experiments. *Frontiers in Human Neuroscience*, *6.* 10.3389/fnhum.2012.00298

Tikka, P. (2008). *Enactive cinema. Simulatorium Eisensteinense*. Publication Series of the University of Art and Design, Helsinki.

Tikka, P., Vuori, R., & Kaipainen, M. (2006). Narrative logic of enactive cinema: *Obsession*. *Digital Creativity, 17*(4), 205–212. 10.1080/14626260601074078

Tikka, P. (2006). Cinema as externalization of consciousness. In M. Punt and R. Pepperell (Eds.), *Screen consciousness: Mind, cinema and world* (pp. 139–162). Rodopi.

Tikka, P. (2005). Dynamic emotion ecology in cinema. In M. Ylä-Kotola & S. Inkinen (Eds.), *Integrated media machine* (pp. 103–126). Lapland University Press.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory*, *30*(3), 167–186. 10.1177/0735275112457914

Varela, F. J., Thompson, E., & Rosch, E. (1991). *Embodied mind: Cognitive science and human experience*. MIT Press.

Vollberg, M. C., & Cikara, M. (2018). The neuroscience of intergroup emotion. *Current Opinion in Psychology, 24*, 48–52. 10.1016/j.copsyc.2018.05.003

Wang, Y., Li, Y., Xiao, W., Fu, Y., & Jie, J. (2020). Investigation on the rationality of the extant ways of scoring the Interpersonal Reactivity Index based on confirmatory factor analysis. *Frontiers in Psychology*, *11.* 10.3389/fpsyg.2020.01086

Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi, 36*, 365–375. 10.1007/s11245-017-9484-6

Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. MIT Press.

Zillmann, D. (1991). Empathy: Affect from bearing witness to the emotions of others. In J. Bryant & D. Zillmann (Eds.), *Responding to the screen: Representation and reaction processes*. Lawrence Erlbaum Associates.

Downloaded from http://direct.mit.edu/pvar/article-pdf/doi/10.1162/pres_a_00378/2138681/pres_a_00378.pdf by Institute of Scientific & Technical Information of China, Shen Dan on 18 September 2023