ObrelaJOURNAL Dialogues in Clinical Neuroscience & Mental Health

e-ISSN: 2585-2795 • p-ISSN: 2654-1432
DOI : 10.26386/obrela.v9i3.398
p. 93-97

The Proteus Effect by Proxy: Neuroethical Implications of Avatar Embodiment, Therapeutic Identity Change and Digital Grief

Nikolaos Statharakos

Adult Psychiatry, General Hospital of Thessaloniki "G. Papanikolaou", Greece

Abstract

Background: Immersive virtual reality (VR) is increasingly moving beyond its traditional role as an exposure-based therapeutic environment and is emerging as a technology capable of modulating bodily self-consciousness, self-perception, and identity. This shift is closely linked to avatar embodiment and the Proteus Effect, whereby users unconsciously adapt their behaviours, attitudes, and self-representations according to the characteristics of their digital avatars. Although these mechanisms may offer promising therapeutic applications in psychiatry, they also raise important neuroethical questions, particularly when applied to vulnerable clinical contexts such as depression, body image disturbance, and grief.

Methods: This hypothesis driven article provides a conceptual synthesis of literature on VR embodiment, the Proteus Effect, therapeutic identity change, digital bereavement, continuing bonds, prolonged grief disorder, and neuroethics. The aim is to examine avatar embodiment as a hypothetical micro-model of therapeutic identity modification and to introduce a novel theoretical framework for understanding immersive interactions with avatars of deceased loved ones.

Results: The paper proposes the concept of the "Proteus Effect by Proxy" to describe a process in which the bereaved individual does not embody the deceased, but interacts with a highly immersive and responsive avatar of the deceased. Through simulated reciprocity, conversational responsiveness, and activation of relational schemas, the avatar may influence the mourner's emotional state, behaviour, autobiographical memory, and post-loss identity. This framework highlights three major neuroethical concerns: threats to patient autonomy and narrative authenticity through implicit identity modification; risks of internalizing algorithmic distortions, biases, or maladaptive relational scripts; and the ethical boundaries of digitally mediated interactions with the deceased, particularly in individuals vulnerable to prolonged grief disorder.

Conclusions: Avatar embodiment represents a powerful frontier in psychiatric technology, with the potential to facilitate therapeutic identity change but also to disrupt autonomy, memory, and grief adaptation. The Proteus Effect by Proxy offers a heuristic framework for examining how immersive digital bereavement technologies may reshape the mourner's identity through interaction with simulated representations of the deceased. Before such technologies are integrated into clinical practice, rigorous ethical, regulatory, and empirical safeguards are required.

Keywords

proteus effect, grief, avatar embodiment, neuroethics, virtual reality, identity

Corresponding Author: Nikolaos Statharakos, Adult Psychiatry, General Hospital of Thessaloniki "G. Papanikolaou", Thessaloniki, Greece, nicstath@gmail.com, nstathaa@auth.gr

Background

Virtual reality (VR) has been utilized experimentally in psychiatry as an advanced medium in exposure therapy, allowing patients to safely confront phobic stimuli or traumatic triggers. (1) However, recent advances in the neuroscience of embodied cognition are shifting this paradigm, presenting novel potentials of VR as not only a mere simulator of external environments but also as a technology capable of modulating the internal architecture of self. (2-4) This shift is anchored by two interconnected phenomena the avatar embodiment and the Proteus effect. Avatar embodiment relies on synchronous multisensory stimulation to induce illusions of body ownership and agency over a virtual body. (2) From a neuroscientific perspective, VR acts as an embodied technology that effectively deceives the brain's predictive coding mechanisms and its body matrix. (3) The Proteus effect describes the psychological phenomenon where individuals unconsciously alter their behaviors and attitudes to conform to the physical and social cues of their avatars. (4-6) Meta-analytic evidence confirms this as a reliable phenomenon driven by combination of self-perception theory, schema activation and deindividuation (7)

By treating these phenomena not as behavioural quirk of novel technologies, but as a mechanistic micro-model of identity change, clinicians can theoretically instantiate candidate identities, elicit identity-congruent behaviors and consolidate those experiences into a patient's real-world self-concept. (8,9) While this offers revolutionary pathways for treating disorders of the self, treating avatars as potent identity-shaping stimulus pushes clinical practice into an uncharted neuroethical territory.

Recent scholarship highlights how VR can uniquely alter the phenomenology of grief by facilitating hyper realistic, embodied continuing bonds with the avatars of the deceased (10), raising a radically new frontier and point of view in the trajectory of grief. Building on this, the paper explores Proteus effect not only as a mechanism for interacting with the lost loved ones but also for internalizing their traits and transforming the bereaved's own identity after the loss. The bereaved enters VR as themselves, but their interaction with the deceased's avatar potentially activates old relational schemas, shifting the bereaved's identity and behaviors. The proxy is the deceased's avatar acting as a catalyst for the mourner's identity shift.

Embodiment as Identity Modification

VR is increasingly framed as an embodied technology that can alter self-representation, not just elicit exposure to feared stimuli**.** (3) The efficacy of avatar embodiment rests on its capacity to modify identity by bypassing entrenched cognitive schemas and supplying experience-near evidence of alternative self-states. (11) Drawing on brain's own simulation architecture, immersive VR enables the systematic manipulation of visceral, motor and sensory reenactments, thereby altering bodily self-consciousness and self-attribution in ways that talk based interventions struggle to reach. (3,12) Embodied interventions have already been used to modulate self-related distortions in pain, eating disorders, depression, and offender populations, by inducing ownership over modified bodies and observing changes in emotion, empathy, and body image. (9) Emerging research demonstrates the potency of this dynamic across multiple psychiatric conditions. (13)

In depression, avatar embodiment can directly reshape internal dialogue. Falconer et al. demonstrated that highly self-critical healthy women who first embodied an adult avatar delivering compassion to a crying virtual child and then embodied the child to receive that same compassion, exhibited significant reductions in self-criticism. This was extended in an open trial with 15 patients meeting criteria for major depressive disorder; three repetitions of an 8-minute VR self-compassion scenario produced significant reductions in depressive symptoms and self-criticism, with PHQ-9 scores decreasing from a mean of 17.54 at baseline to 11.69 at a 4-week follow-up, and four patients achieving clinically significant improvement. (14) Osimo et al. advanced this logic using an embodied self-dialogue paradigm: 58 participants alternated between a virtual body resembling themselves and an avatar of Sigmund Freud. Synchronous body ownership in both roles yielded significantly greater perceived therapeutic change than a non-embodied scripted comparison. (15)

In disorders of the bodily self, VR-based embodiment illusions can systematically overwrite distorted self-representations. A recent systematic review of embodiment illusions in body image disturbance (BID) found that 83.33% of studies reported improvements in BID indices, with medium-to-large effect sizes for both perceptual and affective components. Similarly, in body dysmorphic disorder, rubber-hand-illusion experiments demonstrate proprioceptive drift even under asynchronous stimulation, suggesting aberrant weighting of visual input and perceptual instability rather than a standard multisensory ownership deficit. (13)

Across these contexts, the Proteus Effect functions as a powerful therapeutic catalyst, instantiating an adaptive, avatar-based identity allows patients to generate new episodic evidence that is incompatible with their illness-engulfed self-schemas, opening space for more flexible autobiographical narratives. (6)

However, a critical neuroethical threshold is crossed when this mechanism is exported from controlled psychiatric interventions into the commercial digital afterlife industry. While clinical embodiment aims to constructively overwrite the self, digital bereavement utilizes the same immersive architecture to simulate the absent other. As we will argue, deploying a technology inherently designed to alter identity within the fragile context of grief processing transforms the Proteus Effect from a therapeutic catalyst into a vector for profound psychological disruption.

Digital Bereavement and Proteus Effect by Proxy

The boundary-pushing nature of avatar embodiment is perhaps most evident in the emerging intersection of VR and grief therapy. (10) Standard bereavement models conceptualize adaptive grief through meaning reconstruction (16) and the maintenance of Continuing Bonds (17) which is the process of transforming an external relationship with the deceased into an internalized,

psychological presence. To further explore this intersection a novel theoretical framework termed the "Proteus effect by proxy" is proposed. To understand this, we must explicitly contrast it with standard VR interventions. In conditions like depression, the therapeutic aim is a self-directed Proteus effect: installing a new self-representation by having the patient embody an adaptive avatar. (14,15) In bereavement, however, the phenomenon is other-directed and driven fundamentally by interaction. The bereaved does not embody the deceased, but rather engages in a highly immersive, reciprocal exchange with their avatar. (10)

The Proteus effect occurs by proxy through these interaction dynamics. It is not merely the visual fidelity of the avatar, but the illusion of simulated reciprocity, the conversational turn-taking, programmed micro-expressions, and algorithmic responsiveness that acts as the catalyst. This reciprocal interaction automatically activates deep-seated relational schemas, unconsciously shifting the bereaved's own behaviors, emotional states, and identity back into the pre-loss roles they once occupied (e.g., reverting to "the one cared for by X").

While this simulated interaction could theoretically offer an appealing dialogical space to resolve unfinished business, it is ultimately structurally misaligned with the teleology of grief. Contemporary bereavement models converge on the idea that adaptive mourning involves reconstructing meaning and reorganizing identity around the physical absence and internal psychological presence of the deceased, so that the relationship is transformed and integrated rather than behaviorally maintained as if death had not occurred. (18,19) When a VR encounter with a grief-bot or high-fidelity avatar becomes highly reciprocal, contingent, and on-demand, it weakens the reality of death as a non-negotiable constraint within which the mourner's narrative must be reauthored. Instead of scaffolding a viable post-loss self-story, the interaction risks re-instantiating the premorbid relational script in behavioral form, effectively freezing the bereaved in an outdated identity position organized around the *ongoing* external presence of the deceased. Over time, this may amplify separation distress and complicate meaning-making, particularly when strong continuing bonds are not balanced by the survivor's capacity to make sense of the loss and assimilate it into a progressive sense of identity.

The Ethical Frontiers

Positioning avatar embodiment as a tool for therapeutic identity change and digital grief processing generates profound ethical friction. If VR can reliably rewrite self-perception, the technology ceases to function merely as a therapeutic environment and begins to operate as a psychological prosthetic: a device that does not simply simulate experience but actively restructures the user's identity architecture. This transition from medium to mechanism necessitates rigorous neuroethical oversight that current clinical frameworks do not provide.

The foremost neuroethical challenge concerns the authenticity of therapeutically induced identity change and its relationship to patient autonomy. Clinical avatar embodiment explicitly targets the movement from a stigmatized or maladaptive identity toward a more positive, integrated self. However, the mechanisms through which the Proteus effect achieves this shift remain theoretically contested and, critically, often operate at or below the margins of conscious awareness. The current literature identifies self-perception processes, priming and schema activation, and deindividuation as the leading candidate mechanisms, though none has been definitively isolated, and more recent work has proposed additional explanations including cognitive dissonance, embodiment, and perspective-taking (7). If a patient's self-concept is altered because they unconsciously conformed to the programmed physical characteristics or algorithmically generated responses of an avatar, questions arise regarding the narrative authenticity of that new identity. Does therapeutic avatar embodiment, when it operates primarily through implicit pathways, constitute a form of psychological engineering rather than assisted self-discovery?

This question has direct implications for informed consent. Standard consent procedures address somatic risks such as motion sickness, vestibular disorientation, or transient emotional distress. But identity-modifying interventions demand a fundamentally different consent framework. As Witt argues in the context of neuromodulation, the standard conception of informed consent is seriously lacking because it misses important complexities going along with decisions involving identity changes (20). The problem is structural: a patient consenting from their pre-intervention identity cannot fully evaluate the intervention's desirability from the perspective of the post-intervention identity that does not yet exist. Consenting to what one may later reject, is especially acute when the intervention operates through mechanisms that are themselves difficult to introspect upon. Consent for VR-based identity therapies must therefore evolve beyond acknowledging risks; it must include explicit disclosure that the therapy is designed to modify core self-evaluations and implicit identity schemas through pathways that may bypass conscious deliberation, and it should incorporate perspective-sensitive safeguards that help patients anticipate the experiential consequences of identity change before they undergo it.

The stakes of this ethical challenge escalate sharply when the Proteus effect is applied by proxy to digital bereavement. Where the identity interventions described above target the patient's own self-concept, digital grief processing introduces a second identity into the therapeutic equation, that of the deceased. Ethically, recreating the dead requires navigating their posthumous digital dignity, a concept grounded in the principle that digital remains should be treated with respect analogous to human remains and must not be exploited . Yet the immediate clinical danger lies less with the dignity of the deceased than with the psychological integrity of the survivor.

In populations suffering from, or vulnerable to, Prolonged Grief Disorder (PGD), identity disruption is not an incidental symptom but a core diagnostic feature. The DSM-5-TR lists identity disruption as one of its eight criterion C symptoms, and longitudinal research demonstrates that a post-bereavement merged identity with the deceased is an independent predictor of prolonged grief symptom maintenance, over and above attachment style, relationship type, and initial grief severity (21,22). In this popu-

lation, unstructured interaction with an avatar of the deceased risks acting as a technologically sanctioned mechanism of experiential avoidance that substitutes the simulated presence of the dead for the arduous work of reorganizing identity around their absence. Rather than facilitating adaptive mourning, the avatar may reinforce yearning and separation distress, sustain the fantasy of the deceased's continued external accessibility, or encourage reunion beliefs inconsistent with the reality of death. Empirical work has already linked experiential avoidance to the development and maintenance of PGD, particularly when approach and avoidance behaviors co-occur (23); a compelling VR avatar has the potential to industrialize this co-occurrence.

Equally concerning, Lindemann has argued that deathbot-mediated bonds constitute continuing pseudo-bonds that are qualitatively different from the internalized continuing bonds that contemporary bereavement theory treats as adaptive. (24) Healthy continuing bonds involve the gradual internalization of the deceased's presence, an inner, symbolically maintained relationship that is not dependent on external stimuli. (17,18) When that bond is externalized through a technology platform, when the bereaved comes to depend on an avatar to feel connected to the deceased, the bond becomes fragile, commercially exploitable, and vulnerable to secondary loss if the service is discontinued or degraded. (24) The bereaved's autonomy is correspondingly diminished: they cannot grieve independently of the platform.

A further and perhaps more consequential concern involves the fidelity of the deceased's digital reconstruction. Current AI systems cannot reproduce the unique micro-contingencies, paralinguistic features, relational rhythms, and linguistic idiosyncrasies that constitute a person's interactional identity. No matter how extensive the training data, an AI-generated avatar remains an algorithmic approximation not a veridical reproduction. This limitation intersects dangerously with the very neurocognitive mechanisms that make VR therapeutically powerful. In depression and body-image disorders, avatar embodiment leverages the brain's predictive-processing architecture: VR delivers compelling, multisensory evidence that is discrepant with maladaptive internal models, prompting rapid updating toward healthier self-representations (3). The speed and depth of this updating is a therapeutic asset when the new evidence is clinically designed and accurate. In grief avatar encounters, however, the same updating machinery becomes a profound liability because the "evidence" is algorithmically fabricated. As the immersive avatar recruits the mourner's relational schema, activating stored representations of the deceased's voice, gestures, and interpersonal style, subtle misattunements between the avatar's behavior and the deceased's actual characteristics accumulate as discrepant data points about who the deceased "really was."

Over repeated sessions, particularly in individuals vulnerable to memory modification effects, these dynamic risks a clinically pernicious outcome: the gradual shifting of authentic autobiographical memory toward a synthetic, algorithmically generated approximation. VR experiences are already documented to produce memory effects that blur the boundary between virtual and real events (25); in the affectively charged context of grief, where the mourner is highly motivated to believe in the avatar's fidelity, such effects may be amplified. The risk is compounded by the fact that memory modification, once it has occurred, may be functionally irreversible since the mourner may have no way to recover the original memory trace that has been overwritten by repeated exposure to algorithmically generated approximations.

Clinicians and developers must confront the reality that deploying inaccurate digital resurrections actively revise grief-related models in the wrong direction. It distorts the internalized image of the deceased that continuing bonds theory treats as central to healthy adaptation (17, 18), replacing an authentic internal representation with a technologically scaffolded fiction. Where healthy grief transforms the bond from external attachment to internalized symbolic presence, algorithmically mediated grief risks the opposite trajectory: rewriting the internal representation to match an external artifact that was never accurate in the first place.

These converging risks point to the urgent need for regulatory and professional frameworks that do not yet exist. At a minimum, VR-based grief avatar therapy should be subject to constraints. First, classification as a regulated intervention. Lindemann has argued that deathbots should be classified as medical devices, requiring proof of non-harm and efficacy before deployment. (24) This argument extends forcefully to immersive VR grief avatars, which carry higher risks than text-based chatbots by virtue of their embodied, multisensory nature. Second, clinician competence standards. The interdisciplinary demands of VR grief therapy exceed the training of most mental health professionals. Administering such interventions without specific competence risks exacerbating the very pathologies they aim to address. (26) Third, empirical evidence requirements. At present, no controlled studies examine the effects of VR-based deceased avatar interaction on PGD outcomes, autobiographical memory integrity, or long-term identity coherence. Deploying such technology in the absence of evidence inverts the precautionary principle. Fourth, dual-use governance. The same mechanisms that enable therapeutic identity change through avatar embodiment could, without oversight, be deployed for commercial exploitation of grief, political identity manipulation, or compulsive re-engagement driven by the platform's engagement metrics rather than the patient's clinical needs. (27) The affective potency of embodied VR demands that therapeutic applications be firewalled from commercial imperatives. The ethical frontier of avatar-mediated therapy is, ultimately, the frontier of permissible identity intervention. The question is not whether VR can change who people are but under what conditions, with what safeguards, and with what epistemic honesty about its limitations, such changes can be pursued without violating the autonomy, memory, and psychological integrity of the people they are designed to serve.

Conclusion

The translation of the Proteus effect from cyberpsychological curiosity to clinical intervention represents a watershed moment in psychiatric technology. By leveraging avatar embodiment,

clinicians possess an unprecedented mechanism to bypass cognitive resistance and facilitate therapeutic identity change in depression and body image disorders. However, the power to rapidly alter identity constructs via digital embodiment outpaces our current ethical frameworks. A Proteus-informed clinical protocol must be strictly bound by neuroethical guidelines that safeguard narrative authenticity, prevent the internalization of algorithmic stereotypes, and approach digital bereavement with profound caution. As psychiatry increasingly embraces the metaverse, the discipline must recognize that it is no longer just treating the mind; it is actively re-architecting the embodied self.

Statements

Funding: no funding was received for this work.
Conflict of interest: The author declares no conflict of interest

References

1. Boeldt D, McMahon E, McFaul M, Greenleaf W. Using Virtual Reality Exposure Therapy to Enhance Treatment of Anxiety Disorders: Identifying Areas of Clinical Adoption and Potential Obstacles. Front Psychiatry 2019, 10:773, doi: 10.3389/fpsyt.2019.00773
2. Aymerich-Franch L. Avatar embodiment experiences to enhance mental health. In: Kim J, Song H (eds) Technology and health: Promoting attitude and behavior change. Academic Press, 2020, pp 49–66, doi: 10.1016/B978-0-12-816958-2.00003-4
3. Riva G, Wiederhold BK, Mantovani F. Neuroscience of virtual reality: From virtual exposure to embodied medicine. Cyberpsychol Behav Soc Netw 2019, 22:82–96, doi: 10.1089/cyber.2017.29099.gr
4. Yee N, Bailenson J. The Proteus effect: The effect of transformed self-representation on behavior. Hum Commun Res 2007, 33:271–290, doi: 10.1111/j.1468-2958.2007.00299.x
5. Yee N, Bailenson JN, Ducheneaut N. The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. Commun Res 2009, 36:285–312, doi: 10.1177/0093650208330254
6. Martin Coesel A, Biancardi B, Buisine S. A theoretical review of the Proteus effect: understanding the underlying processes. Front Psychol 2024, 15:1379599, doi: 10.3389/fpsyg.2024.1379599
7. Ratan R, Beyea D, Li BJ, Graciano L. Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the Proteus effect. Media Psychol 2020, 23:651–675, doi: 10.1080/15213269.2019.1623698
8. Vasudha V, Choudhary S. Virtual reality and embodiment: Rethinking self-construction and identity in therapeutic contexts. Front Virtual Real 2026, 7:1756137, doi: 10.3389/frvir.2026.1756137
9. Matamala-Gomez M, Maselli A, Malighetti C, Realdon O, Mantovani F, Riva G. Virtual Body Ownership Illusions for Mental Health: A Narrative Review. J Clin Med 2021, 10:139, doi: 10.3390/jcm10010139
10. Fanti Rovetta F, Valentini D. Grief and virtual reality: Continuing bonds with virtual avatars. Phenomenol Cogn Sci 2025, doi: 10.1007/s11097-025-10072-9
11. Beaufils K, Berland A. Avatar embodiment: From cognitive self-representation to digital body ownership. Hybrid 2022, 9, doi: 10.4000/hybrid.2664
12. Owens EA, Duncan RO. Evidence of a hierarchical representation in bodily self-consciousness: the neural correlates of embodiment and presence in virtual worlds. Front Hum Neurosci 2025, 19:1468947, doi: 10.3389/fnhum.2025.1468947
13. Turbyne C, Goedhart A, de Koning P, Schirmbeck F, Denys D. Systematic Review and Meta-Analysis of Virtual Reality in Mental Healthcare: Effects of Full Body Illusions on Body Image Disturbance. Front Virtual Real 2021, 2:657638, doi: 10.3389/frvir.2021.657638
14. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016, 2:74–80, doi: 10.1192/bjpo.bp.115.002147
15. Osimo SA, Pizarro R, Spanlang B, Slater M. Conversations between self and self as Sigmund Freud--A virtual body ownership paradigm for self counselling. Sci Rep 2015, 5:13899, doi: 10.1038/srep13899
16. Neimeyer RA (ed). Meaning reconstruction & the experience of loss. American Psychological Association, 2001, doi: 10.1037/10397-000
17. Klass D, Silverman PR, Nickman S (eds). Continuing bonds: New understandings of grief. Taylor & Francis, 1996, doi: 10.4324/9781315800790
18. Neimeyer RA, Baldwin SA, Gillies J. Continuing bonds and reconstructing meaning: mitigating complications in bereavement. Death Stud 2006, 30:715–738, doi: 10.1080/07481180600848322
19. Shardlow J. Temporal Perspectives and the Phenomenology of Grief. Rev Philos Psychol 2024, 15:461–482, doi: 10.1007/s13164-022-00659-5
20. Witt K. Identity change and informed consent. J Med Ethics 2017, 43:384–390, doi: 10.1136/medethics-2016-103684
21. Prigerson HG, Boelen PA, Xu J, Smith KV, Maciejewski PK. Validation of the new DSM-5-TR criteria for prolonged grief disorder and the PG-13-Revised (PG-13-R) scale. World Psychiatry 2021, 20:96-106, doi: 10.1002/wps.20823
22. Harris CB, Brookman R, O'Connor M. It's not who you lose, it's who you are: Identity and symptom trajectory in prolonged grief. Curr Psychol 2023, 42:11223–11233, doi: 10.1007/s12144-021-02343-w
23. Eisma MC, Lenferink LIM. Co-occurrence of approach and avoidance in prolonged grief: a latent class analysis. Eur J Psychotraumatol 2023, 14:2190544, doi: 10.1080/20008066.2023.2190544
24. Lindemann NF. The Ethics of 'Deathbots'. Sci Eng Ethics 2022, 28:60, doi: 10.1007/s11948-022-00417-x
25. Johnsdorf M, Pham KA, Schmidt T, Truong VL, Wohnig A, Kisker J et al. New is not always better: Virtual reality does not necessarily enhance mnemonic processing. Front Psychol 2023, 14:1089725, doi: 10.3389/fpsyg.2023.1089725
26. Goulet M-H, Dellazizzo L, Goyer S, Dollé S, Hudon A, Phraxayavong K et al. What Are the Ethical Issues Surrounding Extended Reality in Mental Health? A Scoping Review of the Different Perspectives. Behav Sci 2025, 15:1431, doi: 10.3390/bs15101431
27. Bojić L, Matthes J, Samala AD, Čabarkapa M. The dual impact of virtual reality: examining the addictive potential and therapeutic applications of immersive media in the metaverse. Inform Commun Soc 2026, 29:513–545, doi: 10.1080/1369118X.2025.2520005