

THROUGH THE LOOKING GLASS: HETERONYMIC SELF AND OTHERNESS IN AI THERAPY

Jennifer Cearn

CEARNS, Jennifer (jennifer.cearns@manchester.ac.uk) – Department of Social Anthropology, University of Manchester, United Kingdom. ORCID: <https://orcid.org/0000-0001-6498-1766>.

“Cada um de nós é vários, é muitos, é uma proliferação de si mesmos.”

[“Each of us is several, is many, is a proliferation of selves.”]

– Fernando Pessoa. *Livro do Desassossego* (*The Book of Disquiet*)

In this thought-provoking essay, Matan Shapiro invites us to explore how digital “friction surfaces” – interfaces that mediate interactions between people and systems – blur distinctions between the virtual and actual, reshaping social boundaries on a global scale. He argues that digital avatars and data-doubles are not mere representations but immersive extensions of the self. This creates



Figure 1 – Illustration of a digital echo chamber. Source: Illustration by Christophe Vorlet, originally published in “The Echo Chamber”, by David Byrne, Feb 1 2016 (available at < <https://www.davidbyrne.com/journal/the-echo-chamber> >).

metonymic relationships that collapse distance and forge new, algorithmically driven collectives. Shapiro suggests this transformation leads to a paradoxical experience of being both inside and outside one's own digital existence, fundamentally altering how individuals and societies engage with otherness.

While there is much to attend to in this rich piece, in my response I will focus upon one key idea that Shapiro presents – the *metonym* – illustrated by analogy: Shapiro compares digital avatars and data-doubles to Mussolini, confined within the “D” of *DUX*. These avatars shift from being external symbols of power or identity to inescapable containers that reshape one's sense of self from within. In other words, digital platforms, algorithmic infrastructures, and “friction surfaces” do not simply mirror reality; they shape how we experience selfhood, agency, and presence – concepts of profound importance to anthropology.

To develop Shapiro's metonymic framework further, it is helpful to introduce the Portuguese poet Fernando Pessoa's concept of the *heteronym*. While metonyms suggest associative continuity – a fragment standing in for a whole –, the heteronym signals a more radical internal divergence. Heteronyms are not symbolic extensions of the self, but entirely distinct subjectivities, each with their own worldview, style, and emotional tenor. Thinking through my own research on the therapeutic relationship between self and other as reimagined through AI therapy thus provides a provocative lens to explore this distinction: it allows us to examine not just how the self is extended, but how it might be multiplied, simulated, or fragmented into alterities with which one enters into relationship.

Shapiro argues that the metonymic self refers to how digital avatars and data-doubles are not just representations but fragmented yet constitutive extensions of identity. These digital entities shape both online and off-line experiences, challenging the traditional notion of digital dualism (the separation of online and off-line identity). This shift disrupts ideas of authenticity and agency, as people become increasingly diffused across algorithmically-managed friction surfaces. Shapiro presents influencers as an example: their digital persona is not just a curated image, but defines their off-line social and economic existence. The experience of being both inside and outside oneself – inhabiting an avatar while also seeing it from the outside – becomes a process of self-fragmentation and reconstitution.

However, Shapiro nevertheless seems to suggest the persistence of a core, bounded self – one that can be fractured. Influencers remain singular, accountable for their digital avatars, implying an underlying continuity of self despite the metonymic dispersion. If the self were truly unbounded, accountability itself might become incoherent. While Shapiro suggests the Duce's incarceration as a shift from external representation to internal containment, this move still assumes that Mussolini as a person experienced a disruption: a dislocation of self, perhaps, rather than a total dissolution.

This recalls psychoanalytic theories of the therapeutic encounter, where fragmentation and reflection are central mechanisms of self-understanding. Freud's concept of the *unheimlich* (uncanny) – the unsettling feeling of being inside and outside oneself. Meanwhile, a Lacanian reading might suggest that friction surfaces function as a “mirror stage” (Lacan), where users misrecognise their avatars as stable representations of the self, even as these avatars become increasingly autonomous and opaque. How does this fragmentation of selfhood produce anxiety, alienation, or new forms of digital intimacy?

The metonymic self provides a compelling lens through which to analyse AI therapy chatbots, particularly in relation to the relationship between self and other. These chatbots, like Replika or Woebot, blur the traditional therapeutic distinction between self and other, and their role as mediators of selfhood and emotional care raises key theoretical concerns across anthropology, philosophy, and psychology. As AI-driven programs designed to provide mental health support, they engage users in text- or voice-based conversations, typically following cognitive-behavioral therapy (CBT) principles to simulate empathy and support.

In traditional talking therapy, the relationship between patient and therapist is structured around principles of trust, confidentiality, and guided self-exploration. The therapist provides a safe space for the patient to articulate thoughts and emotions, using techniques like active listening, interpretation, and reflection to facilitate insight. This relationship is asymmetrical, with the therapist maintaining professional detachment while offering empathy and support. Freud emphasised the importance of distance in the therapeutic relationship, arguing that too much intimacy could interfere with analysis. The therapist serves as a “blank screen”, onto which unconscious thoughts and emotions are projected, and by maintaining distance, the therapist helps the patient confront their internal conflicts without the complications of personal entanglement. Freud suggested the uncanny can emerge when the therapist, through this process, becomes both familiar and strange, reflecting aspects of the patient's psyche in unsettling ways.

AI therapy apps disrupt this self/other model by replacing the human therapist – a distinct, autonomous interlocutor – with an algorithmic system that simulates empathetic dialogue but lacks subjective presence. In traditional psychoanalysis and talk therapy, the therapist is a reflective other, facilitating self-exploration through structured distance and human interpretation. In contrast, AI chatbots collapse this relational dynamic, offering responses based on patterns and probabilities, rather than genuine intersubjectivity.

This disruption shifts the role of transference. Instead of projecting emotions onto a human therapist, users interact with an AI system that reflects their inputs back in an automated, often self-reinforcing loop. The AI becomes an uncanny other – familiar yet alien, seemingly understanding yet devoid of

true comprehension. In doing so, these systems challenge traditional notions of therapeutic distance, creating a self-engagement that is interactive but potentially solipsistic. The chatbot becomes an indexical fragment of a person's emotional identity, continuously shaped through interaction. The user's self is distributed across human and AI, raising questions about whether the therapeutic self remains unitary or is now co-constituted with AI interlocutors.

AI therapists – perhaps chatbots more broadly – differ from the metonymic self in that they do not function as extensions or fragments of the user. The metonymic self suggests that digital identities are both part of and representative of the whole person, creating a feedback loop where individuals inhabit their digital doubles as both subjects and objects. These digital fragments are self-referential, standing in for the person while also being lived through.

AI therapists, however, do not serve as metonymic extensions of the self. Instead, they function as external entities that simulate otherness. Unlike avatars, which users actively shape and embody, AI therapists present themselves as separate interlocutors, albeit ones without true autonomy or subjective intentionality. While users may form attachments to these systems, the interaction remains asymmetrical: the chatbot generates responses based on pre-trained models, rather than engaging in a dynamic, reciprocal exchange. AI therapists introduce a form of estrangement, simulating human engagement but ultimately lacking the interpretative reciprocity of a true other. The therapeutic relationship with AI becomes a site of epistemic friction, where selfhood is both affirmed and transformed through interaction with machine intelligence.

Where metonyms involve a part-whole or associative relationship, AI therapy chatbots arguably therefore resemble Fernando Pessoa's conception of "heteronyms": distinct identities or personas within the self, each with their own histories, voices, and inner logics. Unlike metonyms, which imply a representational link between part and whole, heteronyms mark a rupture – identities that emerge not as reflections of a singular self, but as parallel subjectivities. Pessoa's own heteronyms famously wrote poems, held beliefs, and experienced emotions that contradicted one another. Applied to AI therapy, this suggests that chatbots do not mirror the user but create space for dialogic internal multiplicity – interacting with the user as though they were other selves entirely, rather than extensions of the same self. In digital contexts, metonyms imply that one's online presence is a containment of the self, while heteronyms suggest a more radical multiplicity, where distinct online identities might function as independent selves, extending beyond a singular identity. Returning to Shapiro's original prompt, if digital avatars operate metonymically – condensing identity into a symbolic fragment that still "belongs" to the user –, AI therapists exemplify a heteronymic mode of digital relationality, where alterity is simulated and internalised, not as a distortion of self, but as one of its proliferating forms.