ISSN 2280-7853

PHENOMENOLOGY AND MIND

2025 | 29

Rosenberg & Sellier

Topics
Phenomenology and Social Ontology; Ethics and Political Theory; Cognitive Neurosciences, Philosophy of Mind and Language, Logic; Aesthetics, Metaphysics and History of Ideas; Gender Philosophies; European Culture and Politics.

Frequency
2 issues per year

Registrazione del Tribunale di Pavia, n° 6 del 9/07/2012
Direttore responsabile: Roberta De Monticelli
ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

n. 29 - 2025

PHENOMENOLOGY AND MIND

EXPLORING PERSONAL IDENTITY. PHILOSOPHICAL PERSPECTIVES AND INSIGHTS FROM THE ARTS

Edited by Francesca Cesarano, Marco di Feo, Eleonora Volta

Rosenberg & Sellier

Phenomenology and Mind practices double blind refereeing and publishes in English.

CC 2025 Rosenberg & Sellier
via Carlo Alberto 55
10123 Torino
www.rosenbergesellier.it

Rosenberg & Sellier è un marchio registrato utilizzato per concessione della società Traumann s.s.

Phenomenology and Mind, on-line:
http://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

Diego Marconi (Università di Torino)
Gianvito Martino (Università Vita-Salute San Raffaele)
Cristina Meini (Università del Piemonte Orientale)
Andrea Moro (Institute for Advanced Study, IUSS, Pavia)
Elisa Paganini (Università di Milano)
Alfredo Paternoster (Università di Bergamo)
Marco Santambrogio (Università di Parma)
Andrea Sereni (Institute for Advanced Study, IUSS Pavia)
Vera Tripodi (Università di Milano)
Achille Varzi (Columbia University)
Alberto Voltolini (Università di Torino)

History of Ideas (CRISI)
Claudia Baracchi (Università di Milano-Bicocca)
Simonetta Bassi (Università di Pisa)
Andrea Bellantone (Institut Catholique de Toulouse)
Giovanni Bonacina (Università di Urbino)
Enrico Cerasi (Università Vita-Salute San Raffaele)
Francesca Crasta (Università di Cagliari)
Stefano Cristante (Università di Lecce)
Amina Crisma (Università Alma Mater di Bologna)
Giulio D'Onofrio (Università di Salerno)
Catherine Douzou (Université François Rabelais de Tours)
Nicola Gardini (University of Oxford)
Sebastano Ghisu (Università di Sassari)
Simona Langella (Università di Genova)
Anna Marmodoro (University of Oxford)
Vesa Oittinen (University of Helsinki)
Gaetano Rametta (Università di Padova)
Vallori Rasini (Università di Modena e Reggio Emilia)
Francesca Rigotti (Università della Svizzera Italiana)
Hans Bernard Schmid (Universität Basel)
Homero Silveira Santiago (USP – Universidade de São Paulo)
Leonel Ribeiro dos Santos (Universidade de Lisboa)
Alexandre Guimarães Tadeu de Soares (Universidade Federal de Uberlândia, Brasil)
Attilio Scuderi (Università di Catania)
Emidio Spinelli (Università La Sapienza-Roma)
Pirmin Stekeler-Weithofer (Universität Leipzig)
Cristina Terrile (Université François Rabelais de Tours)
Italo Testa (Università di Parma)
Frieder Otto Wolf (Freie Universität Berlin)
Günter Zöller (Ludwig-Maximilians-Universität München)

Phenomenology of gendered personal identity, Gender and Political Normativity, Language and Gender, Philosophy of gender medicine, Women Philosophers (GENDER)
Saray Ayala (California State University, Sacramento)
Maddalena Bonelli (Università di Bergamo)
Antonio Calcagno (King's University College)
Chiara Cappelletto (Università di Milano)
Cristina Colombo (Università Vita-Salute San Raffaele)
Marilisa D'Amico (Università di Milano)
Valentina Di Mattei (Università Vita-Salute San Raffaele)
Luna Donezal (University of Exeter)
Massimo Filippi (Università Vita-Salute San Raffaele)
Sara Heinämaa (University of Jyväskylä)
Dan López De Sa (Universitat de Barcelona)
Anna Loretoni (Scuola Superiore Sant'Anna, Pisa)
Marina Sbisà (Università di Trieste)
Sara Cohen Shabot (University of Haifa)
Alessandra Tanesini (Cardiff University)
Ingrid Vendrell Ferran (Goethe-Universität Frankfurt)
Lea Ypi (London School of Economics)

CONTENTS

CONTENTS

INTRODUCTION

INTRODUCTION

Francesca Cesarano, Marco di Feo, Eleonora Volta
Exploring Personal Identity. Philosophical Perspectives and Insights from the Arts

FRANCESCA CESARANO
Goethe University Frankfurt
Cesarano@em.uni-frankfurt.de

ELEONORA VOLTA
NOVA University of Lisbon
evolta@fcsh.unl.pt

MARCO DI FEO
Vita-Salute San Raffaele University, Milan
mardifco@gmail.com

EXPLORING PERSONAL IDENTITY. PHILOSOPHICAL PERSPECTIVES AND INSIGHTS FROM THE ARTS

Phenomenology and Mind, n. 29 - 2025, pp. 12-21
DOI: 10.17454/pam-2900
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Personal identity has long stood at the crossroads of some of philosophy's most vexed questions. What makes a person the same over time? What constitutes the unity of selfhood amid the flux of experience, the transformations of the body, and the shifting social worlds in which each of us is embedded? Philosophy has approached these questions from multiple angles—metaphysical, epistemological, ethical, and aesthetic. Yet new and pressing questions continue to emerge, particularly in light of contemporary challenges to the very notion of personhood: the rise of artificial intelligence, the politicization of identity, the fragility of memory and embodiment, and the pluralization of gender and social belonging.

This special issue of *Phenomenology and Mind* gathers a selection of contributions that engage with this multifaceted debate. The aim is not to provide a unified or definitive theory of personal identity, but rather to illuminate its constitutive tensions and its manifold developments. The collection brings together phenomenological analyses of subjectivity and embodiment; investigations into the relationship between self and experience, as well as mind and brain, in the process of perceiving one's own identity and personal agency; feminist analyses of gender identity and gendered language; and, finally, aesthetic approaches that view art as both expressing and reshaping identity. By integrating these different perspectives, what emerges is a dynamic framework of analysis that connects self-awareness, affectivity, political action, and creative expression.

1. Rethinking Personhood: From Substance to Relation

Contemporary philosophy has definitely shifted its focus from an isolated, interiorized ego to the being-in-the-world of a lived, embodied subject. The self is no longer conceived as a fixed essence, but as an open horizon of meaning constituted through experience, language, and social interaction. This shift marks a move from what a person is to how personal identity is constituted. The phenomenological approach emphasizes that personal subjectivity is constituted first of all through first-person lived experience: bodily, emotional, practical, and evaluative. The person appears as a unity of sense (*Sinneseinheit*) grounded in intentional life, yet never reducible to a mere psychological continuity or a metaphysical substrate. Through affectivity and value-experience, as Max Scheler[1] and Edith Stein[2] suggested, one's

1 See: M. Scheler, *Formalism in Ethics and Non-Formal Ethics of Values: A New Attempt toward the Foundation of an Ethical Personalism*. Translated by Manfred S. Frings and Roger L. Funk. Evanston, IL: Northwestern University Press, 1973; Id., *The Nature of Sympathy*. Translated by Peter Heath. New Brunswick, NJ: Transaction Publishers (now Routledge), 2008; Id., *The Human Place in the Cosmos*. Translated by Manfred S. Frings. Evanston, IL: Northwestern University Press, 2009.
2 See: E. Stein, *On the Problem of Empathy*. Translated by Waltraut Stein. 3rd revised edition. Washington, D.C.: ICS

individuality takes form not only through reflection but through the very acts of loving, valuing, and willing that express one's "personal center." In this sense, the inquiry into personal identity is inseparable from the study of how we relate to the world and to others. Personal identity thus emerges in an embodied dialectic between self-constitution and intersubjective recognition, between interiority and exposure, constancy and change.

Edoardo Fugali defends the thesis that both self-awareness and world-awareness are inherently embodied and argues that the first-person perspective (FPP) is grounded in the body's role as the zero-point of an oriented spatial field. Selfhood thus arises from the embodied "I can," not from conceptual reflection. Against representational and linguistic accounts that reduce the FPP to ownership of experiences or to conceptual self-reference, Fugali shows that it originates from the bodily structure of perception and action. Through a detailed dialogue with De Vignemont, Baker, and Zahavi, the paper argues that bodily ownership and agency are not independent of the FPP but rather presuppose it. The lived body provides the fundamental frame through which the subject perceives and acts, generating a pre-reflective self-awareness that is prior to any epistemic or linguistic act. Critiquing intellectualist models that define selfhood in terms of higher-order thought, Fugali reinstates embodiment as the primordial locus of first-person givenness. His phenomenological approach reveals that the FPP is not an abstract mental stance but a structural feature of the lived body's sensorimotor intentionality.

Sofia Livi proposes an original contribution to rethinking personhood by grounding selfhood in the lived and sensory dynamics of the body. Moving beyond the narrativist view of the self as a reflective or linguistic construct, Livi examines how respiration and olfaction, two often overlooked modalities, participate in the pre-reflective constitution of personal identity. Through the notion of posture, she articulates a phenomenology of the body that unites the literal configuration of movement with the affective and cognitive stance by which subjects inhabit the world. Breathing and smelling are not just physiological functions, but expressive dimensions of the self, because they shape the way we perceive, feel, and orient ourselves spatially and emotionally. Livi's analysis highlights that perception is never anonymous, but it bears the imprint of an embodied style, a distinctive way of being in the world. By integrating proprioception, affectivity, and atmosphere, she exemplifies how personal identity unfolds in the relational field between body and world.

Stavros Panayiotou presents a profound ethical reconsideration of what it means to be a person, challenging the criterialist view dominant in analytic metaphysics. Criterialism defines personhood through a specific set of psychological or behavioral properties, rationality, self-consciousness, volition, and communication, that are both necessary and sufficient for being a person. Drawing on Emmanuel Levinas, Panayiotou argues that such ontological definitions reduce the person to an object of classification and thereby overlook the ethical foundation of selfhood. For Levinas, personhood does not arise from possessing mental or moral capacities but from the asymmetrical relation of infinite responsibility toward the other. Panayiotou's reading of Levinas defines personhood as relational and non-reciprocal: the self is constituted through responsibility rather than autonomy, through exposure to alterity rather than reflexive self-knowledge. The author thus situates ethics before ontology: the human person is not primarily a rational substance, but an ethical event, an opening brought about by the

Publications, 1989; Ead., *Philosophy of Psychology and the Humanities*. Translated by Mary Catharine Baseheart and Marianne Sawicki. Washington, D.C.: ICS Publications, 2000; Ead., *The Complete Works* – Critical English Edition, vol. 12. Washington, D.C.: ICS Publications / The Catholic University of America Press, 2024.

encounter with the other. This encounter takes on the characteristics of a calling, a taking of a stand that grounds every possible articulation of identity.

2. Metaphysical Questions about Mind, Brain, and Self

Within analytic and philosophy of mind traditions, the problem of personal identity has often been cast in terms of its persistence across time and change and its relation to experience. Competing models, as psychological continuity, bodily continuity, narrative identity, animalism, and constitution theory, have sought to capture the conditions of personal persistence, each revealing different aspects of what "being the same" entails. At the same time, investigations on personhood and the relationship between the self and experience continue to be at the center of an ongoing debate that intersects philosophy of mind, phenomenology and even bioethics, as exemplified by some of the interdisciplinary contributions to this section.

Andrea Bottani challenges one of the central dichotomies structuring contemporary debates on personal identity—the opposition between reductionist ("complex") and anti-reductionist ("simple") views of persons. While this contrast has long been regarded as decisive for understanding the metaphysics of the self, Bottani argues that its significance has been overestimated. He maintains that, although the distinction between reductionist and anti-reductionist frameworks may hold meta-metaphysical relevance, it does little to illuminate the substantive nature of persons or the conditions of their persistence through time. By disentangling questions of personal identity from this entrenched opposition, Bottani invites a reassessment of what truly matters in metaphysical accounts of the self, shifting the focus from abstract structural debates to the substantive questions about the nature and persistence of persons themselves.

Alberto Barbieri offers a rigorous critique of one of the dominant paradigms in contemporary philosophy of mind: the state self-awareness view (SSV), according to which the subject's inner awareness of experience is grounded in the experience's awareness of itself. The author argues that all versions of this theory fail to convincingly explain how a relation at the level of mental states can generate a subject-level phenomenon, such as for-me-ness. The so-called "problem of state awareness" reveals a fundamental explanatory gap: if consciousness is defined by the subject's being aware of its own states, it cannot be reduced to those states being aware of themselves. The paper reorients the discussion toward a view that treats inner awareness not as a reflexive property of mental states but as a primitive relation entertained by the subject. In doing so, it bridges analytic philosophy and phenomenology, reframing the question of personal identity in terms of the lived immediacy of self-presence rather than internal self-representation.

Alfredo Tomasetta reopens the classical Indian debate between Buddhist and Nyāya philosophers on the existence of the self, focusing on the multimodality argument. According to Nyāya thinkers, the very possibility of multimodal perception, where data from distinct sensory modalities (e.g., touch and sight) combine into a single experience, requires a unifying self-capable of synthesizing disparate perceptual inputs. The author examines Monima Chadha's recent attempt, in *Selfless Minds* (2023), to defend Vasubandhu's Buddhist no-self position against this challenge. Chadha claims that multimodal experiences can occur without invoking a substantial self, since sensory data may combine autonomously. Tomasetta argues that this defense fails on two counts: first, it contradicts Vasubandhu's own ontology, which embraces mereological nihilism and momentariness—denying any composite entities like multimodal perceptions; second, it results in an unstable position, oscillating between accepting mental wholes (MMPs) and denying personal unity. Through a detailed analysis of Vasubandhu's metaphysics and the logical implications of Chadha's reply, Tomasetta shows

that the Buddhist no-self view cannot coherently admit multimodality without reintroducing a minimal notion of self.

Luca Zanetti invites a radical rethinking of the very relationship between self and experience. He challenges the common assumption that we are both subjects and bearers of our experiences, emphasizing a crucial distinction between the two. To be the subject of experience, Zanetti argues, is a phenomenological claim about how experience presents itself. By contrast, to be the bearer or owner of experience is a metaphysical claim concerning the relationship between self and experience. Therefore, Zanetti argues that while we undeniably appear as subjects of experience, this does not entail that we are also its bearers. If experience itself constitutes the arena in which the self manifests, then the self occupies no privileged position in it; rather, it is, in a sense, internal to experience. Experience, in this view, "owns" the self, challenging a deeply ingrained assumption in both phenomenology and the philosophy of mind.

Federico Zilio investigates how metaphysical theories of personhood inform ethical attitudes toward patients with disorders of consciousness. After clarifying key notions such as moral status, vegetative state, and metaphysical versus moral personhood, he examines five main theoretical frameworks — personism, animalism, the disjunctive or hybrid view, the constitution view, and ontological personalism — and their respective bioethical implications. Zilio also argues that *ontological personalism* provides a more adequate account for liminal cases, grounding identity in a corporeal entity of rational nature while recognizing profound alterations in psychological manifestation. From a pragmatic neuroethical perspective, Zilio proposes adopting an ontology that attributes intrinsic moral status to patients in vegetative states, thereby ensuring moral protection even amid diagnostic uncertainty.

3. Gender, Intersectionality, and the Reconstruction of Identity

Feminist philosophy and gender studies offer crucial insights into the social constitution of identity. They demonstrate that gendered categories like "man" or "woman" are social products of cultural and political nature. Yet these categories also carry affective and existential weight: they can be at the same time sources of oppression and belonging, constraints and resources for self-formation. Feminist and intersectional perspectives thus invite us to think of personal identity as a site of struggle and creativity, within a continuous negotiation between self-definition and social determination. They challenge philosophy itself to reflect on its own categories, showing that conceptual work is never neutral but always situated within networks of power. The contributions included under this thematic lens investigate how gender and intersectional identities reshape traditional assumptions about personhood and urge us to think about the effects that language has on marginalized identities.

Esa Díaz-León investigates the epistemological foundations of first-person authority in gender self-identifications, engaging with Talia Bettcher's (2009) influential defense of the principle of First-Person Authority (FPA) in gender avowals. Building on recent debates about how best to articulate and justify FPA in ways that support trans rights, Díaz-León proposes an account of the epistemic privilege involved in knowing one's own gender identity. She argues that such self-knowledge partly stems from our introspective access to the conscious manifestations of our dispositional mental states, such as beliefs and desires. By emphasizing the intimate connection between dispositions and their manifestations, Díaz-León shows how this framework can explain why first-person gender avowals carry a distinctive epistemic authority–one grounded in the very nature of self-knowledge and the phenomenology of mental life.

Martina Giovine examines the psychological and ethical implications of misgendering, understood as a form of microaggression that undermines the gender identity of trans*

individuals. Drawing on psychological, medical, and philosophical literature, Giovine explores how structural injustices and social gender norms contribute to the vulnerability of trans* people and affect their identity formation. Against Dembroff and Wodak (2018), who advocate for a radical duty to eliminate gender-specific pronouns as a strategy to address misgendering, Giovine argues instead that gendered language can be essential in affirming personal identity. Her analysis highlights the need for context-sensitive strategies to prevent misgendering and promote inclusive linguistic practices, demonstrating how attention to language can play a pivotal role in supporting selfhood and social recognition.

4. Art, Aesthetic Experience, and the Performance of Selfhood

Finally, art offers a privileged field for exploring personal identity, both as a form of expression, transformation and even resistance. Artistic creation and reception provide a concrete laboratory in which the self is articulated, questioned, and reshaped through gesture, form, and sensibility. The aesthetic dimension discloses aspects of subjectivity that elude abstract theorization, making visible the dynamic interplay between embodiment, imagination, and the social world. In this sense, art not only reflects philosophical inquiry but also enacts it: every artistic act becomes an experiment in selfhood, a testing ground for new ontological and relational possibilities. Through performances, installations, and multimedia practices, artists explore the boundaries between ontological necessity and creative possibility, between personal and collective identity, between authenticity and representation, between human and artificial.

Chiara Cappelletto investigates how the equation of personhood with brain activity—the idea that to "be" means to "have a brain"—has been propagated across visual and performance art, cognitive science, and neuroscience. She argues that brain scans in particular, functioning as "inner portraits", have legitimized a reductive view of the self that neglects the body's entanglement with environment and technology. Analyzing works by artists including Abdoulaye Konaté, Jan Fabre, Pierre Huyghe, and Refik Anadol, Cappelletto challenges neuroimaging's claims to transparency through the lens of visibilization—the process of making visible what resists visualization—and advocates for understanding human subjectivity as situated within intersecting material, epistemic, and narrative networks.

Lisa Sanguineti explores how contemporary artistic practices address questions of identity from decolonial and intersectional perspectives. She examines the ways in which artists from historically marginalized communities use art as a medium of resistance, self-assertion, and identity reappropriation. Through the case study of the 60th Venice Biennale, Sanguineti highlights how diverse artistic experiences engage in processes of identity reclamation, seeking to dismantle stereotypes perpetuated by dominant cultures. Her analysis shows that, for many artists, aesthetic practice becomes both a space of political struggle and a means of reconstructing subjectivity beyond imposed representational frameworks.

Diletta Caimmi offers a historical and philosophical investigation of collective artistic practices and their impact on the conception of identity in contemporary art. Tracing the evolution of group experiments from the technological avant-gardes of the 1960s to the politically engaged collectives of the 1970s, the acronym-based collaborations of the 1980s, and the networked forms of the 1990s, she examines how these shared modes of creation have contributed to deconstructing the romantic myth of the solitary author. Caimmi argues that such collaborative processes give rise to new models of subjectivation that transcend individual boundaries and reflect the complex interrelations among technology, politics, and social interaction in today's artistic landscape.

5. Conclusion

Taken together, the contributions to this special issue converge on an integrative understanding of personal identity as a multidimensional phenomenon: at once

phenomenological and metaphysical, ethical and political, social and aesthetic. Each domain illuminates aspects that the others cannot exhaust. Such integration suggests that the complexity of personal identity demands a plural methodology, within an ongoing dialogue between traditions and disciplines. The philosophical challenge is to articulate this plurality. Moreover, in confronting these questions, the authors gathered here invite us to rethink not only what personal identity is, but why it matters. Exploring personal identity means engaging in a reflection that goes far beyond mere existential concern or theoretical ambition. It means reflecting more broadly on the meaning and value of human life within our societies, in relation to the challenges of technological progress, our professional commitments, our networks of relationships, our responsibilities, and our irrepressible need to express our personal *haecceitas* to its full potential.

SECTION

SECTION 1

RETHINKING PERSONHOOD: FROM SUBSTANCE TO RELATION

EDOARDO FUGALI
University of Messina
edoardoaugusto.fugali@unime.it

EMBODIED FIRST-PERSON PERSPECTIVE

abstract

In this paper I will argue that self-awareness and world-awareness are inherently embodied, with the embodied first-person perspective as their primary structure. Challenging dominant views, I contend that the first-person perspective is not reducible to subjective ownership of experiences or linguistic self-reference but arises from the body's role as the zero-point of an oriented spatial field. This structuring feature of perception and action grounds the cognitive agent's interaction with the environment. The study examines the relationship between first-person perspective and bodily self-awareness, particularly the interplay of ownership and agency, critiquing accounts by De Vignemont, Baker, and Zahavi. It argues that bodily ownership is rooted in the first-person perspective rather than independent of it. Drawing on phenomenology, the paper emphasizes that first-person experience is an intrinsic, pre-reflective structure of embodiment, irreducible to epistemic judgments or conceptual ascriptions, thus challenging reductionist views and reaffirming the primacy of the lived body in self-awareness.

keywords

embodied first-person perspective, bodily self-awareness, sense of ownership and agency, phenomenology of perception, embodied cognition

Phenomenology and Mind, n. 29 - 2025, pp. 22-30
DOI: 10.17454/pam-2903
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In this contribution I uphold the thesis that our awareness of ourselves and of the world around us is intrinsically embodied, at least at its basic level. Embodied first-person perspective serves as the indispensable and primary component of this awareness. To elucidate this claim, a revised definition of first-person perspective is necessary, one that diverges from the dominant approaches proposed in the current literature. First-person perspective is not primarily the awareness of being the owner of subjective experiences and mental states, such as a qualitative feeling of mineness which should attest their belonging to their bearer, or a higher-order conceptual and linguistic capacity which presupposes, among other things, the mastery of the use of the pronoun "I". The expression "first-person perspective" must rather be taken literally, with an explicit reference to the perceptual field of a given cognitive agent as the first, basic occurrence of this phenomenon. Having a perspective is a structural feature inherent to any perceiving being, whose sentient and moving body allows her to disclose an oriented topological space of coordinates, namely a surrounding environment which extends proportionally to her powers of action and perception. As shown by Husserl's analysis of the constitution of space based on the role played by a subject which is essentially and not only contingently embodied, the lived body acts as the absolute zero-point of this space. Conversely, all objects located in it are relative to the position occupied from time to time by the cognitive agent's body and acquire relevance depending on the relationship of proximity or distance to it and as targets of possible or actual actions.

2. Bodily self-awareness: beyond subjective ownership?

A critical task for such an account lies in assessing the link between first-person perspective and sense of bodily self-awareness. The concept of bodily self-awareness needs further clarification due to its multifaceted nature. A key distinction must be drawn between the sense of bodily ownership and the sense of agency (Gallagher, 2000). Bodily ownership consists in the sense of belongingness of our body and its parts to ourselves, while agency encompasses the awareness of being the initiator of one's own actions. Both features contribute, to varying degrees, to the production and refinement of bodily self-awareness. While bodily ownership manifests as a continuous and passive sense of embodiment, persisting during stillness and involuntary movements, agency emerges solely during active engagement in purposeful and intentional actions. Phenomenologically, the sense of ownership and the sense of agency appear intricately interwoven in our everyday interactions with the world, often blurring into a seemingly indistinguishable whole. Only when we pause our actions do we become aware of our body as an inert and voluminous inner space, vaguely delimited by the boundaries of

our skin, yet distinctly separate from the external world. Both the sense of ownership and the sense of agency operate in a pre-reflective and non-conceptual manner. Our body is not a prominent object of intentional focus; rather, it is "lived through" during action, receding into the background as our attention fixates on the objects we manipulate. It is only when we actively refrain from action, allowing our body to rest, that we can thematically shift our attention inwards and turn it towards the body itself.

Regarding the question of the connection between bodily awareness and first-person perspective, it is worth considering De Vignemont's account. According to her, first-person component is detachable from the experience of ownership, since the latter can occur without an explicit awareness of it. This suggests that the first-person perspective properly arises at the level of judgment, i.e., an explicit, conceptual self-attribution of body parts and mental states, which presupposes the mastery of the pronoun "I" and therefore linguistic capacities (De Vignemont, 2007, p. 444). However, separating the first-person component from bodily ownership leads to a rather counterintuitive consequence, namely postulating an anonymous level of bodily representations where the body is not yet qualified as "one's own". Moreover, De Vignemont (2018, pp. 49-64) seems to conflate the first-personal character of bodily self-awareness with the phenomenon of immunity to error through misidentification relative to the first person (IEM), analysed by S. Shoemaker. IEM denotes the *epistemic* infallibility of immediate self-ascription of both mental and bodily states to oneself, and as such, is a property of *judgments* rather than a feature of a phenomenological experience, let alone a perceptual one.[1] As Shoemaker puts it, self-awareness does not imply being presented to oneself as an object and therefore cannot be assimilated to a kind of perception, either internal or external. When I utter the sentence "I feel a pain", I am pointing to myself as the one feeling the pain, not to the pain as something existing regardless of me and the feeling I am having of it (Shoemaker, 1968, pp. 563-564).

De Vignemont is right in distinguishing the first-person perspective from the subjective feeling of bodily ownership, provided that their difference is only conceptual. However, her account falls short in establishing the foundational relationship between both terms. It is bodily ownership what requires and is grounded in the first-person perspective, not the other way around. It is by virtue of structural constraints, such as the internal provenience of bodily sensations like proprioception and kinaesthesia, or the invariance of the egocentric visuo-spatial frame of reference, that we acknowledge our body as our own. De Vignemont herself mentions Ernst Mach's famous self-portrait, which graphically represents our optical array – a faceless body around which a perspectivally deformed environment spreads out. However, instead of considering this perceptual experience in its unique characteristic, she confines herself to merely underline its role in grounding epistemic IEM-judgments.

3. Rudimentary and robust first-person perspective: a too intellectualistic account

Even L.R. Baker appears to hold a similar intellectualistic view. She defines first-person perspective as "the capacity of thinking of oneself, conceived in the first person, as the object of one's thought" (Baker, 2013, p. XIX). Baker rightly points out the basic fact that first-person perspective is a dispositional property that necessarily belongs to the ontological world array, distinguishing between two layers: rudimentary and robust first-person perspective, corresponding to consciousness and self-consciousness. However, she unilaterally overemphasizes the role played by the latter. Rudimentary first-person perspective does not

1 Even more radically than De Vignemont, J.L. Bermúdez (2017) claims that bodily awareness and first-person perspective arise solely at the cognitive level of IEM-judgments: there is nothing like a qualitative feeling of ownership as a primitive phenomenological feature or, at least, it is devoid of any cognitive and epistemic function.

require fully developed self-concepts or language, but it is rather defined by the essential feature of having a point of view from the spatiotemporal location the organism occupies, from which it perceives and interacts with the world. This level requires only sentience and intentionality, as well as the capacity for goal-directed behavior. In contrast, robust first-person perspective entails the ability to use well-defined concepts and a fully developed language to refer to oneself as oneself from one's own point of view. A typical example of this ability is the use of the pronoun "I" within a complex first-person sentence introduced by a "that"-clause and governed by a psychological main verb.[2] Such sentences express I-thoughts, which consist of self-attributions of a first-person reference and are intertwined with our daily actions in several ways, such as when we are called to account for them by offering justifications and reasons. Thus, both rudimentary and robust first-person perspectives share the property of immunity to error through misidentification, albeit for different reasons: by referring to oneself through "I"-sentences and self-concepts, one never fails to recognize oneself as such, but in the case of rudimentary first-person perspective, no error occurs simply because this capacity is lacking at all (Baker, 2013, pp. 33-43).

Against Baker's view, it can be argued that bodily self-awareness is also immune to errors of misidentification because it constitutes a genuine and positive—albeit pre-reflective and implicit—way of referring to oneself as oneself. My ability to discriminate myself from other objects in my environment is not primarily due to the robust first-person perspective; rather, it derives from my self-awareness as an embodied subject, embedded in an environment and capable of action and perception. Having a first-person perspective entails as its basic condition of possibility the very fact that I am incorporated in a moving body as the center of an oriented space. Therefore, first-person perspective consists in the capacity to perceive the world from the standpoint afforded by one's body. I can distinguish myself from others by virtue of being rooted in my own body, making my perspective inherently non-interchangeable with anyone else's. The robust first-person perspective is grounded in and constituted by the rudimentary one to a greater extent than Baker admits. The only difference is that the robust first-person perspective includes the ability to refer to oneself through concepts. The act of recognizing oneself as oneself occurs entirely at the rudimentary level and does not necessarily involve a fully developed process of explicit self-reflection. While concepts are used at this level, they are best characterized as proto-concepts that do not require fully developed linguistic capacities and acts of predication. There is no substantial difference between pointing to oneself with a gesture and appropriately using the pronoun "I." The robustness of the higher-order stage of first-person perspective depends not so much on the appropriate use of "I"-concepts but rather on the capacity to elaborate complex narratives about oneself through reflection.

Unlike Baker, I do not believe that the rudimentary first-person perspective entirely lacks self-concepts. Even in non-human organisms and infants, consciousness always involves a form of self-consciousness, albeit not reflective in character and not developed through linguistic resources. For an infant or an organism to be self-conscious, only primitive, barely representational, and non-linguistic self-concepts are required. These proto-concepts are acquired through repeated acts of self-recognition of a purely perceptual nature, which imply an indirect intentional reference to one's own body alongside the straightforward

2 Baker refers here to Castañeda's account, which posits that self-consciousness is irreducible to any form of third-person reference as shown by the use of the pronouns "I" and "he*" (i.e. "he, himself"). These expressions denote peculiar logical categories that cannot be replaced either by nouns or by descriptions in which they do not occur. First-person statements hold referential, ontological and epistemological priority over all names and descriptions, implying that every form of object awareness presupposes self-awareness as its foundations (Castañeda, 1966).

intentionality directed toward external objects. Only through these proto-concepts can more sophisticated and refined self-concepts develop, referring to the subject as an embodied being. I can use these self-concepts appropriately only to the extent that I perceive myself unceasingly within a specific environmental context. First-person states do not relate to an abstract self or ego but to whole embodied and embedded personal subjects.

4. From "I" to "It"? Why "I" won't fit into third-person statements

Baker devotes considerable effort to her criticism of the semantics-based accounts of the first-person perspective, supported by scholars such as Kaplan and Perry, by opposing them with well-grounded ontological arguments. According to D. Kaplan (1989), a first-person statement like "I am DK" can be reduced to a third-person statement without any loss of content, as its truth conditions can be expressed without recourse to the pronoun "I". Kaplan argues that there is nothing special distinguishing the use of "I" from other indexical terms. The statements in which it occurs do not provide any information on "further facts" about the irreducible subjectivity of my point of view, as "I" has no fixed, recognizable content and therefore does not denote a self-concept; hence, there are no genuine first-person facts or properties. Yet, as Baker points out, this line of reasoning misses the target. The bare fact of having a first-person perspective and being a person is the intentional correlate of a self-perception that obtains even in the absence of any thought and utterance. "No semantic rule seems adequate fully to explain my being LB" (Baker, 2013, p. XIV). Being me supposes that I am existing like an embodied person, which constitutes a primitive ontological category.[3] My being the person I am is not captured by any statement about it, because I continue to be who I am even in absence of every thought or utterance about myself. It is not primarily the capacity to entertain I-thoughts that ensures the reference of pronoun "I" and other indexicals, but rather the mere fact of being an embodied subject centered on its own perspectival zero-point, always given in a continuous stream of internal perceptions.

A more moderate reductionist approach, such as the one supported by J. Perry (1979; 2002), acknowledges the existence of first-person thoughts expressible in genuine I-sentences that cannot be translated into third-person descriptions. These sentences display beliefs and others cognitive attitudes endowed with a motivational force that causally modifies the speaker's behavior, which is lost once they are converted into third-person sentences. Despite this, Perry argues that these sentences possess at most an epistemic and pragmatic value and do not reveal any special facts about the existence of persons. Self-notions are merely "the normal repository of normally self-informative ways of perceiving and the normal motivator of normally self-dependent ways of acting" (Perry, 2002, p. 202). Therefore, there is no crucial difference between first-person and third-person notions of oneself from an ontological standpoint. Perry's famous example of the messy shopper nicely illustrates this point. While pushing my shopping trolley in a supermarket, I notice someone leaving a trail of sugar from a punctured pack. As I try to find out who is making the mess, a glance in the mirror reveals to me that I am that one shopper. In this case I have two unlinked notions of myself that become unified once I recognize their identity through collecting information about objective facts – i.e. being someone else and being oneself – which are conveyed through different cognitive channels. With this said, the difference between being myself and being someone else is not ontologically unsurmountable; it depends only on the agent-relative role played by

3 In a similar vein, S. Kripke opposes Kaplan's assertion that the pronoun "I" has no fixed meaning and cannot denote a self-concept, claiming that its reference is exhausted by a linguistic rule of use. On the contrary, Kripke argues that the reference of "I" is fixed, though not by means of a concept or representation, but through one's own immediate self-acquaintance. "I" refers to the subject of such awareness, which involves "the whole person in the ordinary sense" (Kripke, 2011, p. 308) including her own body.

the subject who ascertains the identity of both through practical and epistemic procedures. The beliefs "I am JP" and "JP is JP" share the same subject matter content but differ only in their reflective content, that is, the way they are held. Baker objects to this view by asserting that the act of recognition of oneself as oneself precedes the evaluation of information and any (subpersonal) association between self-concepts and third-person concepts. From the very first moment, I must be endowed with a first-personal perspective of myself *as myself*, which alone allows every act of self-recognition (Baker, 2013, pp. 53-54).[4] Hence, the first-person perspective is not the by-product of an epistemic attitude but an irreducible and ineliminable ontological feature of reality, as I am simultaneously both the bearer and the intentional object of every perception and thought about myself.

5. The "same old body" strikes back again: Why "I" can't escape the flesh and the world

Similar considerations concerning the original and irreducible nature of the first-person perspective are developed by D. Zahavi (1999; 2005), who emphasizes the role of embodied experience in its foundation. Like Shoemaker, Perry, Baker and the other philosophers mentioned earlier, Zahavi appeals to the property of IEM which distinguishes first-personal statements containing the pronoun "I". As a pure indexical, "I" unfailingly refers to its denotatum without requiring further description, guaranteeing not only the existence but also the immediate experiential presence of its referent whenever is uttered, without presenting it as an object. According to Zahavi, to be aware of myself as a subject, I do not need positive knowledge about my objective properties, such as being a physical object, a biological organism or a person. Every form of objective knowledge about myself must be grounded in the end on self-awareness, under pain of falling into infinite regress. "My awareness of myself [...] is not mediated by the awareness of any identifying property" (Zahavi, 1999, p. 9). The subject of which I am aware is neither something beyond my experience nor the intentional correlate of an experience. Instead, it simply expresses the first-personal mode through which all experiences are given, making them irreducible to intentional objects.

I disagree with this statement. "I" is not merely an empty reference point, being defined in its properties by the thoughts or mental states occurring here and now, even though it cannot be reduced to them. The necessary and contingent condition for using "I" lies in its connection to a moving and perceiving body. Serving as the zero-point of orientation, the body grounds us in our environment, enabling us to refer to every perceptual object and, conversely, to position ourselves in relation to them. The presence of indexicals in language indicates, prior to any rule of usage, the fundamental fact that I, as an embodied subject, possess a first-person perspective on the world. In every act of perception directed toward a particular object, there is always an implicit reference to myself and to the zero-point where my body is situated. Consequently, the use of demonstratives like "this" and "that" gains its proper function only in relation to a pure and absolute (i.e. "essential") indexicals such as "I", "here" and "now". Understanding the rules governing the use of indexicals presupposes to exist as a bodily subject who simultaneously encounters herself and the worldly context where she perceives and acts. The first-person perspective, as rooted in a moving and perceiving body, is not an

4 Baker's objections can also be raised against C. Peacocke's (2017) "agency-involving account", where he argues that only representations of action can allow the shift from the mere awareness of the presence of an anonymous body to the awareness of this body as mine. According to Peacocke, a set of subpersonal representations of both bodily actions and objective properties, generated by an "integrating apparatus", is necessary for this. However, contrary to Peacocke's assertion, first-person non-conceptual content is inseparable from a basic form of embodied consciousness. Moreover, it seems highly counterintuitive to propose representational first-person contents that cannot be directly experienced.

ineffable feature merged with experiential contents. Despite its pre-reflective nature, it is a structural datum that can be subject to thematic attention.

The first fact that immediately captures our attention is the distinctly different way our body manifests itself compared to any other object, prompting us to question whether it can be considered an object at all. First and foremost, we must discern between perceiving our body in an object-like fashion through exteroceptive sensations and feeling it from within via bodily interoceptive sensory channels. It is only in the former case that the body can be equated to any other external object, whereas, when experienced internally, it constitutes something uniquely its own without parallel. These two ways of experiencing the body correspond to what E. Husserl calls material body, or body as object (*Körper*), and lived body, or body as subject (*Leib*) (Husserl, 1989).

Indeed, only the lived body comes into appearance through bodily sensations such as proprioception, internal passive touch, kinaesthesias, vestibular and nociceptive sensations, and so forth, filling its internal space. This intracorporeal space contrasts with external worldly space due to its relative uniformity and lack of perspectivity and constitutes the central focus from which the latter unfolds, according to the power of action of the lived body. Husserl characterizes the lived body as an "I sense" and an "I can", namely as a twofold but unitary power of sensation and movement, corresponding at least in part to the sense of bodily ownership and the sense of agency. Both aspects belong to a unitary phenomenon. Whereas proprioception and inner tactile sensations give rise to the body sensed from within, kinaesthesias are responsible for the awareness of the body as a center of movement and action and as a source of a functioning intentionality that is directed toward external objects, enabling the deployment of an oriented space which has the lived body itself as its zero-point. The body not only appears from a first-person perspective but also acts as its condition of possibility by virtue of its sensorimotor constitution. The first-person perspective is rooted in the spatial location occupied by the subject due to its embodied constitution and therefore is co-given with the self-manifestation of the body. The metaphor of the ego as a pole of action and affection, from which our intentional acts radiate, acquires its full meaning only by virtue of its reference to a lived body as the center of orientation for perception and action.[5]

However, contrary to what Zahavi claims (1999, p. 104), the lived body too is the intentional term of a perceptual act, albeit of a peculiar way, since it can be accessed from the inside. The fact that the body is the center of my perspective on the world, and it is not comparable to an external object does not prevent me from perceiving it. While the body itself is not given perspectivally, I cannot reduce its entire dimensional field to the absolute "here" in which it is located. The scope of perception is not limited to the body as object, as the self-constitution of the lived body implies its self-manifestation. As far as the first-personal mode of givenness reveals a conscious state belonging to someone with all its qualitative contents, this bearer – especially when considered in her embodied constitution – must be understood as more than an evanescent and not-objectifiable entity. As Zahavi himself admits (2005, pp. 26 and 205), for the lived body to constitute the objective side of bodily self-awareness, it must be an integral part of our common world. As an embodied subject I have a perspective to the world, but this does not imply that I am entirely foreign to it, as claimed by the proponents of

5 According to Alsmith (2017), an embodied subject can develop an awareness of being located in a world of things that afford her actions without needing an explicit representation of herself as the perspectival center of her perceptions. However, the way we interact with objects through our actions requires at least a pre-reflective awareness of the different spatial relationships we have with them, depending on their varying distances from our body.

a supposed "purity" of a transcendental subject that hovers over the world from a perspective fundamentally alien to it.

6. Conclusions

In this study I have argued that the first-person perspective is inherently embodied and irreducible to higher-order cognitive processes or linguistic self-reference. The body, as the zero-point of an oriented spatial field, structures perceptual and motor capacities, enabling self-awareness and interaction with the environment. Through a critique of De Vignemont, Baker, Kaplan, and Perry, the analysis has tried to show that bodily ownership and agency are not independent from, but rather rooted in, the first-person perspective. While conceptual and linguistic capacities can offer an invaluable contribution to refining (bodily) self-awareness, they do not constitute however its fundamental basis. Instead, selfhood is primarily grounded in the pre-reflective lived experience of the body, which remains irreducible both to epistemic and representational accounts relying on the dogma of reducibility to a third-personal "view from nowhere". By reaffirming the primacy of embodiment, this perspective challenges reductionist views and underscores the necessity of a phenomenological approach to understanding first-person experience.

REFERENCES

Alsmith, A.J.T. (2017). Perspectival structure and agentive self-location. In De Vignemont & Alsmith (2017), 263-287;

Baker, L. R. (2013). *Naturalism and the first-person perspective*. Oxford-New York, NY: Oxford University Press;

Bermúdez, J. L. (2017). Ownership and the space of the body. In De Vignemont & Alsmith (2017), 117-144;

Castañeda, H.-N. (1966). 'He': A study in the logic of self-consciousness. *Ratio,* 8, 130-157;

De Vignemont, F. (2007). Habeas Corpus: The sense of ownership of one's own body. *Mind & Language,* 4(22), 427-449. doi:10.1111/j.1468-0017.2007.00315.x;

De Vignemont, F. (2018). *Mind the body. An exploration of bodily self-awareness*. Oxford-New York, NY: Oxford University Press;

De Vignemont, F., & Alsmith, A. J. T. (Eds.) (2017). *The subject's matter. self-consciousness and the body*. Cambridge, MA: MIT Press;

Gallagher, S. (2000): Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, 1(4), 14-21. doi:10.1016/s1364-6613(99)01417-5;

Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Second book. Studies in the phenomenology of constitution*. Trans. R. Rojcewicz & A. Schuwer. Dordrecht: Springer;

Kaplan, D. (1989). Demonstratives. An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, H. K. Wettstein, & D. Kaplan (Eds.), *Themes from Kaplan* (pp. 481-563). Oxford-New York, NY: Oxford University Press;

Kripke, S. (2011). The first person. In Id., *Philosophical troubles. Collected papers* (pp. 292-321). Oxford-New York, NY: Oxford University Press;

Peacocke, C. (2017). Philosophical reflections on the first person, the body, and agency. In De Vignemont & Alsmith (2017), 289-310;

Perry, J. (1979). The problem of the essential indexical. *Noûs* 1(13), 3-21. doi:10.2307/2214792;

Perry, J. (2002). *Identity, personal identity, and the self*. Indianapolis, IN: Hackett;

Shoemaker, S. (1968): Self-reference and self-awareness. *The Journal of Philosophy*, 65(19), 555-567. doi:10.2307/2024121;

Zahavi, D. (1999). *Self-awareness and alterity. A phenomenological investigation*. Evaston, IL: Northwestern University Press;
Zahavi, D. (2005). *Subjectivity and selfhood. Investigating the first-person perspective*. Cambridge, MA: MIT Press.

SOFIA LIVI
Scuola Normale Superiore
sofia.livi@sns.it

BREATHING POSTURES

abstract

This paper examines the relationship between breathing, olfactory experience, and the sense of self. By employing the concept of posture, I explore the proprioceptive dimension of breathing within the complex framework of olfaction. The link between respiration and smell is analyzed phenomenologically, situating it within broader discussions on selfhood and embodiment. Ultimately, this research highlights how breathing and smelling are integral to our posture – our distinctive way of being in the world.

keywords

perception, olfaction, posture, self, phenomenology

Phenomenology and Mind, n. 29 - 2025, pp. 32-45
DOI: 10.17454/pam-2904
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Smell is often described as a fleeting, elusive perceptual modality, slipping away from the focus of consciousness (Sela & Sobel, 2010). Yet, as recent studies suggest, olfaction deeply contributes to our overall phenomenological experience (Smith, 2022) and to our embodied self-awareness (Young, 2023).[1] Following this perspective, I examine the entanglement between respiration, olfactory perception, and selfhood.

The guiding question of this paper aligns with what Schechtman (2007) calls the *characterization question*: what makes you the person that you are, with your peculiar personal identity? While narrativist accounts emphasize self-construction through storytelling, I argue that bodily and perceptual factors – specifically breathing and olfaction – are integral to our self-experience. To support this claim, I introduce the concept of *posture* as a lens through which to understand how our body orients itself in the world, both physically and affectively. The upshot of this research is that perception is not an anonymous matter: the perceptual experience of inhabiting the world is imbued with meaning and profoundly intertwined with the experience of the body.

The structure of the paper is as follows. Firstly, I present the problem of the self I want to explore (§1). I introduce the concept of posture (§2) and apply it in the investigation of breathing (§3). The interrelation between olfaction, the activity of breathing, and spatial exploration is then analyzed with the more general intention of facing the problem of what makes us the persons that we are (§4/6). All the various arguments converge in the last section (§7), in which breathing and the activity of smell are framed as components of our posture – that is to say, of our style of being in the world.

1. The problems of the self

The problem of the self is not a simple, monolithic question.[2] The self is a multifaceted concept analyzable from many perspectives: for instance, one might be to investigate what the self is, as an entity, or examine the conditions of persistence of selves over time. In this paper, I shall be discussing what the qualities are that define the identity of a self. In other words, the

1 "Our sense of self is partially unconsciously constructed from smelling our odourprint and smells of others [...]. [Smell] generates a partial foundation of our continual construction of our embodied and culturally embedded sense of self" (Young, 2023, p. 197).

2 For some introductions to the problems, see Di Francesco (1998), Gallagher (2000).

properties that make you *you*. And, connectedly, the elements that shape the feeling of having a particular identity.

As anticipated, this issue has been labelled as *the characterization question* and investigates "which characteristics are truly those of some person" (Schechtman, 2007, p. 73). One influential answer to the dilemma has been delivered in the solution of narrativism, according to which the characteristics of the self are narrative productions, entanglements of beliefs that subjects have of themselves.[3] Those narrations are story-like beliefs, like the plot of a novel.[4]

While I am not arguing here against the influence of narrations for the self,[5] my intention is to carve a space for other types of factors that might answer the characterization question. In doing so, I share the purpose of theorists[6] who aim at explaining psychological phenomena by appealing to less demanding processes rather than complex, full-fledged beliefs. For instance, in the case of the characterization question, this strategy has been recently advocated for by Colombetti and Bogotá (2024, p. 608): in their account, "we understand ourselves tacitly or pre-reflectively (and thus non-narratively) as embodied and situated". I will explore how the experience of our body, meshed with perceptual activity, is an element in forging the self and the phenomenological feeling of being a certain person with specific psychological orientations, history, and preferences.

2. Postures

I will develop my argument by means of the concept of posture. In its everyday use, this concept has a twofold meaning: one semantic dimension points to the musculoskeletal configuration of an organism in space; the other one is metaphorical and addresses the more general stance that a person takes towards their environment. This second connotation of the word is used to define the way in which subjects face the world: one can say, for instance, that a person holds a progressive posture in political debates, or that their posture is tight and aggressive when engaging in arguments.

This second layer of meaning might be defined as the disposition of a subject towards the world (that could have many nuances, e.g., behavioral, cognitive, and affective), a style of engaging with the environment given certain situations.[7] Asking what kind of posture is typically assumed in various circumstances – e.g., what kind of posture does a person have when immersed in a crowd of strangers – is a way of exploring how subjects develop themselves in time and space.[8]

These dispositional properties are distinct but intertwined with self-narrations, the set of characteristics that are represented (implicitly or explicitly) by the subject themselves (and by other persons too) as being the core of their self-definition. Of course, the way your style of engaging with the world is not impermeable to your self-narrations and vice versa. However, the dispositional properties that subjects have can deviate from the beliefs they have about themselves: beliefs might be open to representational error. Let us take an example. What

3 See MacIntyre (2013), Ricoeur (1992), Schechtman (2007). For a review of the various narrativist positions, Schechtman (2011).

4 According to Schechtman, narratives need not to be explicit.

5 For the limits of such an approach, see Witt (2020).

6 Typically, this is a desideratum of embodied and situated programs of research.

7 The dispositional concept of posture is close to that of *habit* and *habitus.* See Leboeuf (2020) for an introductory sketch of those two concepts (respectively, from the thought of Maurice Merleau-Ponty and Pierre Bourdieu) in the debate over implicit bias.

8 "Our activities and practices carve themselves into our bodies through repetition, shaping our bodies over time in distinctive ways and becoming habits. What we do shapes not just our muscles (e.g., by bulking them up), but also our posture (which depends on the totality of the musculoskeletal system) and, perhaps least obviously of all, the overall *style* of our actions, movements, and expressions" (Colombetti & Bogotá, 2024, p. 610).

if you depicted yourself as being an emotionally strong and mature person, while actually repressing your affective life because of your unconscious tendency to avoid emotions? In asking what characteristics should be attributed to a subject, it might be sensible to investigate their posture, conceived as an affective and intellectual style of engaging with the world, instead of relying on the beliefs that the subjects have about themselves. The self-narrations that one has can display slight distortions, incoherence, or sheer confabulations.[9]

So far, I have analyzed the two meanings of the world *posture* and argued that one semantic pole is of interest in the investigation of the characterization question. But the peculiarity of the word *posture* is that even if the two meanings can be separated, there is a strong relationship between the two. In one sense, if we define the metaphorical posture as the style of engaging with the world in certain situations, then it should also include the way in which subjects typically conduct their bodies in various contexts (their *literal* posture). In another sense, the bodily postural configuration of a subject deeply influences their affective, cognitive postures and vice versa. This ambiguity reflects the entanglement between body and experience: a subject's bodily comportment influences, and is influenced by, their perceptual and affective orientation (Merleau-Ponty, 2012).

Indeed, the way your body is configured in space seems to influence the way you affectively respond to the world: for instance, situations of stress are faced more positively (with a better attitude, self-esteem and mood) if you adopt an upright posture instead of a slumped one (Nair *et al*., 2015). Another study's result highlights how a slouched muscular posture is an obstacle in recovering from bad moods (Veenstra *et al.,* 2017). Musculoskeletal posture also seems to have an impact on evaluating retrieved memories (Michalak *et al*., 2014). This evidence might be interpreted by saying that the bodily posture influences the phenomenal experience and the style of a subjects' responses to the world: if your style of swaying around (i.e., the disposition to move your body in certain contexts) is more chest-open and straight, then in stressful situations you might be disposed to faster emotional recoveries. This idea is in line with another study that reported how posture influences, as well, the feeling of having power over situations (Carney *et al.,* 2010). In sum, the habit of moving the body in a certain fashion can shape your affective posture, for example, making you more emotionally flexible or assertive, and influence the phenomenology of your experience (e.g., for what concerns the feeling of powerfulness).

On the other hand, ideas and affective reactions influence the way in which the body is physically articulated. It has been reported how emotional episodes activate distinct patterns of movement (Dael *et al.,* 2012). If you entertain the disposition to feel in determinate ways in certain settings, you will probably be disposed to move your bodies in specific ways. Thoughts too may have a similar property, as the beliefs (also the implicit ones) over one's own identity and place in the world can influence the way people conduct their bodies. For instance, a phenomenological study by Iris Marion Young (1990) analyses how women tend to move their bodies in a more constricted way. This idea has been sustained by psychological research (de Lemus *et al.,* 2012). In parallel, the embodied life of a black person can be influenced by social encounters and the projected social expectations and beliefs, leading to the adoption of more remissive bodily postures (Yancy, 2008).

The mentioned studies support the idea that the literal and the metaphorical postures are knotted together by recursive influences. Affective, cognitive, and embodied behavioral dispositions influence each other. In the rest of the text, with the concept of *posture* I will point

9 It should be noted that this last remark is not incompatible with some narrativist positions, such as Ricoeur's (1992).

to this complex intertwining, referring to the style of the encounter with the world that also resonates in the way the body is configured in space.

In the attempt to explore a possible nuance to the characterization, I will examine the relationship between posture, breathing, and the sense of smell.

3. Breathing postures

As seen, the way the body moves in space has relevance for the overall definition of a subject's posture. But what exactly is included in the concept of bodily posture? I have suggested how it alludes to the musculoskeletal configuration of a body in space. With this word, we normally indicate the style in which our everyday bodily activity is conducted, for instance, evoking the potential manners in which limbs sway in space, or the possible curvature of the spinal cord. One can think about the style of moving shoulders, the rhythm of the gait, or maybe the tone of our gesticulating habits.

Leaving aside these facets of bodily posture, I slide to an element that is not ordinarily taken to be part of the notion of posture: the activity of breathing. Breathing is the motoric activity that necessarily underlies all other movements. Far from being just an interoceptive matter,[10] it involves our core muscles, even influencing the expansion and contraction of the pelvic floor, and therefore our proprioceptive system.[11] The *proprioceptive system* is the body's ability to sense its position, movement, and orientation in space without relying on visual input (Proske & Gandevia, 2012).[12] The most influential way to account for this kind of awareness is to appeal to spatial representations of the body – for instance, with the concepts of *body schema* (Gallagher, 2006) or *body map* (De Vignemont, 2018).

Breathing relies on the activation of the diaphragm (Fogarty *et al.,* 2018).[13] The diaphragmatic sensations are one form of proprioceptive awareness (at least in our accepted definition). This proprioceptive awareness enables volitional regulation, as the feedback on this muscle (a type of feedback that can be consciously registered) enables the regulation of breathing movements. In other words, to be aware of the tension or relaxation of the diaphragm – i.e., of its movement – enables the subject to consciously control it. The inhalation of air can be volitionally regulated, and this is because the movement of the diaphragm is proprioceptively accessible.

Even if this goes against common sense, breathing should be appreciated as part of our embodied postural configuration, as a component of the musculoskeletal disposition in space. It concerns the dynamics of our body, the rhythm of its movements, and the way it occupies space, playing a role in maintaining balance and stability (Kocjan *et al.,* 2018; Stephens *et al.,* 2017). We should notice how breathing is not a uniform process. Our respiratory activity can

10 The term interoception refers to the perception of one's homeostatic state, for instance, the perception of hunger or thirst. See Craig (2003).

11 If it is true that diaphragmatic muscles are in some sense internal, according to the definition of proprioception and interoception that I adopt here, the perception of their contraction or expansion is a proprioceptive matter. This contraction or expansion does not involve the homeostatic state of the subject, but rather the way they move their body in space. I thank Louise Richardson for raising the issue of whether the perception of the state of the internal muscles might be interoceptive.

12 "Traditionally [...] the term proprioceptor has been restricted to receptors concerned with conscious sensations, and these include the senses of limb position and movement, the sense of tension or force, the sense of effort, and the sense of balance" (Proske & Gandevia, 2012, p. 1651).

13 In this paper, for simplicity, I will use the terms breathing and sniffing as synonyms. However, differences can be appreciated between the two (Xi *et al.,* 2023). If in both cases, the diaphragm plays a central role (Benício *et al.,* 2021), accessory respiratory muscles can also play a role in sniffing activity (Katagiri *et al.,* 2003). While the diaphragm is central to sniffing, accessory muscles can compensate to some extent, especially when diaphragmatic activation is limited or controlled. What matters for my argument is that for effective and efficient sniffing, the diaphragm's involvement (but also, more in general, proprioceptive sensations) remains crucial.

undergo variations: for instance, we can breathe by inflating our chest, with a swift rhythm, or, perhaps, taking deeper diaphragmatic breaths. Similarly to how we straighten or curve our back, our respiration expresses and influences facets of our internal state (for instance, our emotional condition). Indeed, breathing is not just a matter of inhaling and expelling air: the regulation of air intake has a lot to do with the modulation of mood (Perciavalle *et al.*, 2017). This is evident, for example, in cases of anxiety and fear, where the breath seems to slip out of our control, becoming rapid and tense; or in the instance of sighing, a possible expression of emotions such as nostalgia, sadness, or relief.[14]

If the style in which we (consciously or unconsciously) position our chin is part of our bodily posture, then also the way we (consciously or unconsciously) inflate our trunk should be considered as such. For brevity, we shall label a subject's dispositional style of breathing in different contexts as their *breathing posture.*

4. Smelling and breathing

Breathing also has a fundamental role in olfaction. In the literature, sniffing has not always been considered to be a necessary requisite for perceiving smells (see the discussion in Young, 2017), as it has been observed, in laboratory conditions, that olfactory experiences can be elicited without the act of sniffing.[15]

However, breathing and sniffing constantly accompany smell experience, and it is difficult to phenomenologically distinguish the two. Interestingly, Louise Richardson (2013) has noted how the act of sniffing is an essential part of the feeling of exteroceptivity involved in the sense of smell. She writes:

> Olfactory experience is exteroceptive, I shall argue, in that it involves smells seeming to be brought in from without by sniffing. [...] The experience you have is one of air being brought *into* the nostrils, *from without*, though the air is not represented as being at any distance or direction from you. (Richardson, 2013, pp. 402-410)

In Richardson, sniffing is necessary for the exteroceptive component (the perception of something external to the subject) in the perception of smells. However, this exteroceptivity has no spatial qualifications. In other words, while odors are perceived as being external to the subject (thanks to the activity of sniffing), they are just located in a spatially indeterminate area around the nose. She describes odors as seeming to be 'simply here', in a vague spatial proximity without a definable shape or volume.[16]

Let us unpack her argument a little. Richardson's position assigns to olfaction more perceptual complexity than traditionally allowed. Throughout the history of philosophy, smell has been considered to be without any interesting perceptual value.[17] Pure sensations, odors have been said to be felt "in the nose" (Smith, 2005, p. 139) in the same way a headache can be said to be felt in the head. Connecting olfaction to breathing (and in particular to sniffing), Richardson aims at giving a more sophisticated picture of smell perception, including in this last concept the tactile sensation of air passing through our respiratory system. These

14 I thank Louise Richardson for suggesting some examples here.

15 Young discusses the results obtained by Sobel *et al.* (1998). In their experiments, olfactory experience can be elicited without active sniffing by the subject, but by blowing air in their nostrils. Young thus comments: "[S]niffing is not necessary for us to perceive smells and even when sniffing is used to gain access to the olfactory realm of stimuli the motoric component is inessential. Moreover, the somatosensory experience of airflow can be a sufficient condition of undergoing olfactory experiences – even passively" (2017, p. 105).

16 See also Batty (2010).

17 For a brief history of the philosophical positions on smells, see Keller & Young (2023, pp. 1-11).

tactile sensations, according to Richardson, are part of our experience of odors, and they are responsible for the exteroceptive sensation that something external to us is coming into contact with our sensory systems. In this way, we can perceive smells as something outside of us (differently from a headache). Richardson's solution hinges on a reflection on the individuation of the senses: how should we discriminate a tactile experience from an olfactory one?[18] She argues against a conjunctive position on perceptual experience, i.e., the idea that the overall experience of sniffing a scent is the conjunction of two independently specifiable perceptual experiences (an olfactory one plus a tactile one). She argues, instead, that sensory experiences should be considered as part of a unified perceptual system: the olfactory experience (e.g., vanillariness) and the tactile experience (e.g., the experience of the air in the nostrils) are "both abstractions from the overall experience" (e.g., the full experience of sniffing vanillariness) (Richardson, 2013, p. 415).

5. Breathing and the olfactory exploration of space

Following Richardson's remarks about the holistic nature of experience, the point could be integrated by another reflection on the complex, unitary nature of the perception of smells. According to some scientists, the sniff is part of the activity of smelling (Mainland & Sobel, 2006). As it has been argued, "[i]t is now clear that sniffing is part of the olfactory percept" (Ferdenzi et al., 2015, p. 1777); and since in sniffing and breathing not only the tactile sensations of airflow through the nose or mouth are involved, but also the proprioceptive sensation of the muscles of the abdomen expanding and contracting; therefore, the proprioceptive sensations involved in the regulation of breath should be taken into account in the phenomenological experience of encountering smells.[19]

If, as Richardson states, the tactile sensation contributes to the phenomenological exteroceptive component (i.e., the experiential feeling that I am coming into contact with something external to me), on the other side the proprioceptive sensations might contribute, instead, to the feeling that *more or less of this something external to me is explored.*

I now explain why. In the activity of breathing, more or less muscular effort can be put into the inhalation or the expulsion of air. The degree of contraction of the diaphragmatic muscle (and the respective proprioceptive sensation) is an indicator of how much force is enacted in the exploration of the environment. Breathing deeply, engaging the diaphragm in a more intense way, will elicit the proprioceptive sensation that the overall muscles of the trunk enlarge – the chest rises, the belly inflates with air. This proprioceptive sensation might be described as a feeling of bodily expansion. In parallel, this sensation of the broadening of the body is accompanied by the feeling that *more air* is explored.

Indeed, imagine yourself in an olfactory, purposeful exploration of the environment: your instinctive movements would be either to breathe more deeply, or to sniff in a faster manner (and maybe move the head to detect from which direction a smell is coming from).[20] The spatial expansion of the core part of the body could be linked with a greater sense of

18 The question of the identification of the senses often revolves around the philosophical or scientific discussion of how we distinguish and categorize different types of sensory experiences. See, for instance, Fulkerson (2014), Macpherson (2011).

19 Here, my interest is to argue in favor of the inclusion of proprioception in the analysis of olfactory experiences. However, in the holistic phenomenology of sniffing and smelling, other components might be taken into consideration: for example, chemosensory or interoceptive sensations.

20 "Perception of an odour in the environment usually initiates a sniffing episode composed of several sniffs. Each sniff appears to be of shorter duration and to have a greater inhalation velocity than a normal breath, and it is generally accepted that this behaviour enhances perception of an odour [...] [T]his behaviour may enhance odour perception by increasing the amount and rate at which odour molecules reach the olfactory receptor epithelium" (Laing, 1983, pp. 99-102).

amplitude in one's olfactory exploration of the environment. The movements we make with our diaphragm engage our bodies so that the quality of the sense of our being in and exploring space becomes different. The perception of our spatial contact with the world enlarges, as well as the interchanges with the air we breathe. On the contrary, sniffing in a constricted way (for instance, when one is anxious) can trigger, phenomenologically speaking, suffocating sensations and the feeling of not freely sampling the air around. Reduced diaphragmatic movement sustains the feeling that the space of your exploration is limited. Indeed, when one is close to a smell they find disgusting, the instinct is to diminish the strength of their breathing to limit the exposure to the volume of air around them (I will explore this last point in the next section).

In line with the phenomenological tradition, breathing is a *mode of being-in-the-world:* the lived body is the foundation of self-experience. For Husserl (1989), the body is not merely a biological entity among others in the world (*Körper*), but a *Leib*, a lived and pre-reflective medium through which the world is disclosed. In this sense, breathing can be appreciated as a posture: our breathing patterns, together with being objective physiological functions, subjectively participate in the constitution of perceptual experience and self-awareness. Husserl's notion of *kinesthetic consciousness* (Husserl, 1980) further supports this point: kinesthetic sensations – including those involved in respiration – contribute to our sense of spatiality and embodiment. Breathing, in this sense, is not an isolated bodily function but an integral part of the lived body's ongoing engagement with the world.

To sum up the findings of this section: the air we sample with our sniffing is not completely spatially unqualified (as, instead, in Richardson), but is perceived as being more or less vast in connection with the spatial sensations of the body and the movements it engages in. In these cases, the quality and the amount of space explored are indissociable from the quality of inhabiting our bodies. These arguments portray olfaction as a complex activity that also involves the breathing posture and, in parallel, the feeling of exploring space in more expansive or constricted manners.

6. Pleasant smells

I have, so far, explored the importance of sniffing and breathing for smelling. However, the direction of the analysis can also be reversed: sniffing is also subconsciously modulated by the perceived pleasantness of the smells. Various studies have reported that during the sniff, the volume of air intake is regulated according to the perceived pleasantness or unpleasantness of odours (Bensafi *et al.,* 2002, 2003; Ferdenzi *et al.,* 2015; Johnson *et al*., 2003): pleasant odorants are inhaled more deeply and strongly, while those rated as unpleasant are sniffed less vigorously.[21] This behavior is reflected in the typical expression of disgust: the disgusted face shrinks the nostrils, so that the airflow is reduced. Young (2017, p. 101) summarizes these findings: "we modulate our sniff responses in a very robust and fine-grained manner to the valence of an olfactory object. [...] the volume of air intake and strength of motoric inhalation are modulated in accordance with the pleasant or unpleasant nature of the stimulus". Accordingly, Ferdenzi *et al.* (2015, p. 2) write that "there is now psychophysiological evidence that sniffing is modulated not only by odor intensity but also by subjective pleasantness".

The pleasantness that smells have for us is then a factor in the modulation of our breathing posture and the sensation of exploring the environment. But does it mean that the same odors have the same effects on everyone? Do good odors have the power to make us breathe more deeply? The point here is complicated. Together with quality and intensity, hedonics (i.e., pleasantness) is one of the three dimensions of olfactory experience (Rouby & Bensafi, 2002;

21 For a discussion of the ecological meaning of sniffing modulation, see Ferdenzi *et al.* (2015, pp. 7-9).

Wilson & Stevenson, 2006), and is not only dependent on the chemical configuration of the stimulus (Arshamian *et al*., 2022; Khan *et al.,* 2007), but is also modulated by various elements from the sphere of the subject, among which: their homeostatic state (Cabanac, 1971), their culture (Ferdenzi *et al.,* 2013; Kapoor, 2022), their personal history (Herz *et al*., 2004). Also, the multimodal cues presented to the subject at the moment of sniffing are another factor in the determination of the hedonic value of a smell for a given subject.[22]

Since there are many factors that play a role in defining whether a stimulus would be pleasant or unpleasant for a subject, it is misleading to talk about *good odors.* Odors move us, they are attractive or repellent to us, but it is not only a matter of biology or molecular structure.[23] The affective perception of odors (the pleasant or unpleasant perception) is also influenced by cultural and biographical factors. For example, familiarity with a smell is one variability that increases the pleasantness of a smell for a subject (Ferdenzi *et al.,* 2013).

Olfaction exemplifies how perception must be framed beyond biological reductionism – rather, as the historian of the senses Constance Classen argues, perception is (also) a cultural act (Classen, 1993). Cultural norms can regulate what we consider to be repelling odors. The cultural values of social contexts contribute to the definition of what should be perceived as pleasant and offensive. This is coherent with the presence of an olfactory component in social discrimination, such as racism (Kettler, 2020): some smells become unpleasant for subjects because associated with contexts that they despise or deem disgusting. In this sense, experiences of smells seem to be influenced not only by autobiographic memories, but also by beliefs and shared ideas.

Given the importance of experience and culture, olfactory preferences should be considered (at least to a certain degree) learnt.[24] The experiences, thoughts, and beliefs you associate with odors can influence the affective sensations felt when exposed to them (Herz, 2006; Herz *et al.,* 2004). Therefore, for every subject with olfactory abilities, there will be one array of smells experienced as being pleasant or unpleasant. And given the importance of both culture and biographical memories for shaping this olfactory palette, each array of smells classifies and distinguishes the subject both as a member of a certain group and as an individual. The memories, as well as the cultural and social positioning of a subject, constitute a peculiar kaleidoscope that shapes the affective reactions in smell perception.

So, even in this context, it is useful to talk about dispositions to perceive certain smells to be pleasant or unpleasant. To make an example: since I am familiar with the odor of my room and associate it with comfort and a safe personal space, every time that I sniff it, I would be disposed to perceive pleasant sensations. If a stranger entered my room, they would not have, probably, the same disposition, hence the same positive experience. That is because their olfactory affective dispositions, shaped by different experiences and emotional associations, would be different from mine.

7. Breathing and olfactory postures

I am about to knot together all the argumentative threads. So far, I have argued that the analysis of posture can be relevant to the study of the problem of the self (as formulated by the characterization question). I have discussed the tie between literal and metaphoric meanings of posture and explored the concept of breathing posture. This has led to an appreciation of the interplay between breathing and smelling: breathing is the proprioceptive component of

22 "Except for a few notable cases [...] most smells are markedly ambiguous in their hedonic assessments by humans. Ambiguous means that the same stimulus can have a different appeal in separate settings depending on the type of encounter" (Barwich, 2020, p. 130).

23 Even those have of course a certain impact, e.g. see Arshamian *et al*. (2022).

24 For the concept of perceptual learning in olfaction, see Wilson and Stevenson (2006).

the olfactory experience that opens the phenomenological dimension of exploring more or less space. I have reported how breathing can be modulated by the perceived pleasantness of smells, and then I have moved to the problem: what is it that makes an odor pleasant or unpleasant for a subject? The answer has turned out to be complex, including biological as well as cultural and biographical factors.

From these arguments, it is possible to draw some conclusions.

Firstly, it can be argued that associations, past experiences, and the culture of a subject can shape their posture towards odors, the affective dispositions towards them. The olfactory preferences of a subject, their *olfactory posture,* concern the places, people, foods towards which they feel to belonging, and, conversely, about the things they consider to be repellent and alien to them. In this sense, olfactory posture cannot be analyzed without considering the meaning that certain smells and situations have for the individual. Being such a complex fingerprint, the way the subject is touched by odors is an affective characterization of the self: an affective postural style. The olfactory posture of a subject should be taken into account when investigating the characterization question, not only because it relates to their personal history, but also because it affects next encounters – the behavioral, affective, and embodied ways of opening to the world.

Secondly, as we have previously seen, the hedonic value that smells have for us also modulate the way in which we breathe: this points to a regulation not only of our bodily movements but, as we saw, also of our proprioceptive feelings of bodily expansion and contraction, and of our sensation of spatial (olfactive) exploration of the air around. Not only will smelling an aroma that we consider bad elicit a negative judgment in us, but it will also shape our breathing posture and our sense of exploring space. Implicitly, the encounter with odors is impactful in our sense of being immersed in atmospheres.[25] We have dispositions to like or dislike certain odors, and these dispositions can trigger in us, for instance, the sensations to be surrounded by stifling environments: this way, olfactory and embodied postures shape the phenomenology of the subject.[26] For understanding the style of engaging with the environments in various contexts, it is important to unveil how the person is affected by perceptual encounters and how their sensation of being immersed in an environment is nuanced by different tonalities.

The experience of smelling should therefore be considered a multimodal encounter with the world that involves not only our nose and brain, but also the way our body breathes, moves in space. It does not only revolve around the smelling source, but it resonates with past scented encounters, olfactory heritages, ideas, and emotions towards the objects and the situations, shaping our sense of belonging. When we smell something, our perception can be pleasant or unpleasant; it may include a tone of familiarity, of recognition, and closeness. This discloses an olfactory style, a way to be affected by odors. And this perceptual postural configuration is meshed with the embodied one: the breathing patterns, as I have illustrated previously, are susceptible to the influences of smell pleasantness for the subject. As breathing plays roles in the way the body moves in space and in the sensation of exploring olfactorily the external space (as regulated by the activities of sniffing and breathing), those elements are implicitly modified in the encounter with scents and aromas. This continuous interplay

25 See the concepts of bodily contraction and expansion in the perception of atmospheres in the philosophy of Hermann Schmitz. For an introduction: Schmitz *et al.* (2011).

26 The role of olfaction in shaping experience is expressed by Barry Smith: "Olfaction, whether conscious or not, serves to offer a constant background to our conscious lives, modulating emotions, attention, awareness, perception of our surroundings and ourselves, guiding and regulating successful food choice and consolidating episodic memory" (Smith, 2022, p. 32).

characterizes a facet of the perception: the way we perceive the world turns out not to be a neutral, disembodied matter. Rather, through olfaction, organisms are affected and interpret the world.

In sum, I argue that olfactory posture – our embodied and affective dispositional style of engaging with smells – contributes to self-characterization: olfaction, far from being a mere access to sensory properties, is a constitutive dimension of selfhood.

REFERENCES

Arshamian, A., Gerkin, R. C., Kruspe, N., Wnuk, E., Floyd, S., O'Meara, C., Rodriguez, G. G., Lundström, J. N., Mainland, J. D., & Majid, A. (2022). The perception of odor pleasantness is shared across cultures. *Current Biology*, *32*(9), 2061-2066. https://doi.org/10.1016/j.cub.2022.03.014;

Benício, K., Resqueti, V. R., Dias, F. A. L., Pennati, F., Aliverti, A., Medeiros da Fonseca, J. D., & Fregonezi, G. A. F. (2021). Effects of diaphragmatic control on multiparametric analysis of the sniff nasal inspiratory pressure test and inspiratory muscle activity in healthy subjects. *PLoS ONE*, *16*(7), e0253132. https://doi.org/10.1371/journal.pone.0253132;

Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., & Holley, A. (2002). Autonomic nervous system responses to odours: The role of pleasantness and arousal. *Chemical Senses*, *27*(8), 703-709;

Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., & Holley, A. (2003). Perceptual, affective, and cognitive judgments of odors: Pleasantness and handedness effects. *Brain and Cognition*, *51*(3), 270-275;

Cabanac, M. (1971). Physiological role of pleasure: A stimulus can feel pleasant or unpleasant depending upon its usefulness as determined by internal signals. *Science*, *173*(4002), 1103-1107. https://doi.org/10.1126/science.173.4002.1103;

Carney, D. R., Cuddy, A. J. C., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363-1368. https://doi.org/10.1177/0956797610383437;

Classen, C. (1993). *Worlds of sense: Exploring the senses in history and across cultures*. New York, NY: Routledge. https://books.google.com/books?hl=it&lr=&id=mnCpEAAAQBAJ;

Colombetti, G., & Bogotá, J. D. (2024). The tacitly situated self: From narration to sedimentation and projection. *Topoi*, *43*(3), 607-615. https://doi.org/10.1007/s11245-024-10044-9;

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology*, *13*(4), 500-505;

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085-1091;

de Lemus, S., Spears, R., & Moya, M. (2012). The power of a smile to move you: Complementary submissiveness in women's posture as a function of gender salience and facial expression. *Personality and Social Psychology Bulletin*, *38*(11), 1480-1494. https://doi.org/10.1177/0146167212454178;

De Vignemont, F. (2018). *Mind the body: An exploration of bodily self-awareness*. Oxford: Oxford University Press;

Di Francesco, M. (1998). *L'io e i suoi sé. Identità personale e scienza della mente*. Available at: https://research.iusspavia.it/handle/20.500.12076/1855;

Ferdenzi, C., Fournel, A., Thévenet, M., Coppin, G., & Bensafi, M. (2015). Viewing olfactory affective responses through the sniff prism: Effect of perceptual dimensions and age on olfactomotor responses to odors. *Frontiers in Psychology*, *6*, Article 1776. https://doi.org/10.3389/fpsyg.2015.01776;

Ferdenzi, C., Roberts, S. C., Schirmer, A., Delplanque, S., Cekic, S., Porcherot, C., Cayeux, I.,

Sander, D., & Grandjean, D. (2013). Variability of affective responses to odors: Culture, gender, and olfactory knowledge. *Chemical Senses*, *38*(2), 175-186;
Fogarty, M. J., Mantilla, C. B., & Sieck, G. C. (2018). Breathing: Motor control of diaphragm muscle. *Physiology*, *33*(2), 113-126. https://doi.org/10.1152/physiol.00002.2018;
Fulkerson, M. (2014). Rethinking the senses and their interactions: The case for sensory pluralism. *Frontiers in Psychology*, *5*, Article 1426. https://doi.org/10.3389/fpsyg.2014.01426;
Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14-21;
Gallagher, S. (2006). *How the body shapes the mind*. Oxford: Clarendon Press;
Herz, R. S. (2006). I know what I like: Understanding odor preferences. In J. Drobnick (Ed.), *The smell culture reader* (pp. 190-203). New York, NY: Berg;
Herz, R. S., Schankler, C., & Beland, S. (2004). Olfaction, emotion, and associative learning: Effects on motivated behavior. *Motivation and Emotion*, *28*(4), 363-383. https://doi.org/10.1007/s11031-004-2389-x;
Husserl, E. (1980). *Phenomenology and the foundations of the sciences*. The Hague, Netherlands: Nijhoff;
Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy*. Dordrecht, Netherlands: Springer;
Johnson, B. N., Mainland, J. D., & Sobel, N. (2003). Rapid olfactory processing implicates subcortical control of an olfactomotor system. *Journal of Neurophysiology*, *90*(2), 1084-1094. https://doi.org/10.1152/jn.00115.2003;
Kapoor, S. (2022). The smells of caste – Body, self, and politics. In N. Di Stefano & M. T. Russo (Eds.), *Olfaction: An interdisciplinary perspective from philosophy to life sciences* (Vol. 4, pp. 21-34). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-81659-3_3;
Katagiri, M., Abe, T., Yokoba, M., Dobashi, Y., Tomita, T., & Easton, P. A. (2003). Neck and abdominal muscle activity during a sniff. *Respiratory Medicine*, *97*(9), 1027-1035. https://doi.org/10.1016/s0954-6111(03)00133-1;
Keller, A., & Young, B. D. (2023). *Theoretical perspectives on smell*. New York, NY: Routledge. https://doi.org/10.4324/9781003207801;
Kettler, A. (2020). *The smell of slavery: Olfactory racism and the Atlantic world*. Cambridge: Cambridge University Press;
Khan, R. M., Luk, C.-H., Flinker, A., Aggarwal, A., Lapid, H., Haddad, R., & Sobel, N. (2007). Predicting odor pleasantness from odorant structure: Pleasantness as a reflection of the physical world. *Journal of Neuroscience*, *27*(37), 10015-10023;
Kocjan, J., Gzik-Zroska, B., Nowakowska, K., Burkacki, M., Suchoń, S., Michnik, R., Czyżewski, D., & Adamek, M. (2018). Impact of diaphragm function parameters on balance maintenance. *PLoS ONE*, *13*(12), e0208697. https://doi.org/10.1371/journal.pone.0208697;
Laing, D. G. (1983). Natural sniffing gives optimum odour perception for humans. *Perception*, *12*(2), 99-117. https://doi.org/10.1068/p120099;
Leboeuf, C. (2020). The embodied biased mind. In N. Levy & L. Johnston (Eds.), *An introduction to implicit bias* (pp. 41-56). New York, NY: Routledge;
MacIntyre, A. (2013). *After virtue*. London: A&C Black. https://books.google.com/books?hl=it&lr=&id=td_UAAAAQBAJ;
Macpherson, F. (2011). Taxonomising the senses. *Philosophical Studies*, *153*(1), 123-142. https://doi.org/10.1007/s11098-010-9643-8;
Mainland, J., & Sobel, N. (2006). The sniff is part of the olfactory percept. *Chemical Senses*, *31*(2), 181-196;
Merleau-Ponty, M. (2012). *Phenomenology of perception*. London: Routledge;

Michalak, J., Mischnat, J., & Teismann, T. (2014). Sitting posture makes a difference—Embodiment effects on depressive memory bias. *Clinical Psychology & Psychotherapy*, *21*(6), 519-524. https://doi.org/10.1002/cpp.1890;
Nair, S., Sagar, M., Sollers III, J., Consedine, N., & Broadbent, E. (2015). Do slumped and upright postures affect stress responses? A randomized trial. *Health Psychology*, *34*(6), 632-641;
Perciavalle, V., Blandini, M., Fecarotta, P., Buscemi, A., Di Corrado, D., Bertolo, L., Fichera, F., & Coco, M. (2017). The role of deep breathing on stress. *Neurological Sciences*, *38*(3), 451-458. https://doi.org/10.1007/s10072-016-2790-8;
Proske, U., & Gandevia, S. C. (2012). The proprioceptive senses: Their roles in signaling body shape, body position and movement, and muscle force. *Physiological Reviews*, *92*(4), 1651-1697. https://doi.org/10.1152/physrev.00048.2011;
Richardson, L. (2013). Sniffing and smelling. *Philosophical Studies*, *162*(2), 401-419. https://doi.org/10.1007/s11098-011-9774-6;
Ricoeur, P. (1992). *Oneself as another*. Chicago, IL: University of Chicago Press. https://books.google.com/books?hl=it&lr=&id=uCZSOYcB_CIC;
Rouby, C., & Bensafi, M. (2002). Is there a hedonic dimension to odors? In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 140-159). Cambridge: Cambridge University Press;
Schechtman, M. (2007). *The constitution of selves*. Ithaca, NY: Cornell University Press. https://books.google.com/books?hl=it&lr=&id=8NhXyKt7KUwC&oi=fnd&pg=PP11&dq=schechtman&ots=2yqvwGo-1L&sig=CWgMk3CH_bq2cVFl9d4HiIOZyuM;
Schechtman, M. (2011). The narrative self. In S. Gallagher (Ed.), *The Oxford handbook of the self*. Oxford: Oxford University Press. https://academic.oup.com/edited-volume/38581/chapter/334606560;
Schmitz, H., Müllan, R. O., & Slaby, J. (2011). Emotions outside the box—The new phenomenology of feeling and corporeality. *Phenomenology and the Cognitive Sciences*, *10*(2), 241-259. https://doi.org/10.1007/s11097-011-9195-1;
Sela, L., & Sobel, N. (2010). Human olfaction: A constant state of change-blindness. *Experimental Brain Research*, *205*(1), 13-29. https://doi.org/10.1007/s00221-010-2348-6;
Smith, A. D. (2005). *The problem of perception*. Delhi: Motilal Banarsidass Publishers;
Smith, B. C. (2022). The role of smell in consciousness. In B. C. Smith (Ed.), *Theoretical perspectives on smell* (pp. 13-35). Abingdon: Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003207801-3/role-smell-consciousness-barry-smith;
Sobel, N., Prabhakaran, V., Desmond, J. E., Glover, G. H., Goode, R. L., Sullivan, E. V., & Gabrieli, J. D. E. (1998). Sniffing and smelling: Separate subsystems in the human olfactory cortex. *Nature*, *392*(6673), 282-286. https://doi.org/10.1038/32654;
Stephens, R. J., Haas, M., Moore, W. L., Emmil, J. R., Sipress, J. A., & Williams, A. (2017). Effects of diaphragmatic breathing patterns on balance: A preliminary clinical trial. *Journal of Manipulative and Physiological Therapeutics*, *40*(3), 169-175. https://doi.org/10.1016/j.jmpt.2017.01.005;
Veenstra, L., Schneider, I. K., & Koole, S. L. (2017). Embodied mood regulation: The impact of body posture on mood recovery, negative thoughts, and mood-congruent recall. *Cognition and Emotion*, *31*(7), 1361-1376. https://doi.org/10.1080/02699931.2016.1225003;
Wilson, D. A., & Stevenson, R. J. (2006). *Learning to smell: Olfactory perception from neurobiology to behavior*. Baltimore, MD: Johns Hopkins University Press. https://books.google.com/books?hl=it&lr=&id=Fwb3rCJqDJAC&oi=fnd&pg=PP13&dq=learning+to+smell&ots=mxl02VKiiX&sig=1Vlci1QLH_l6i48B_j5WI6lQ928;
Witt, K. (2020). Narrative and characterization. *Erkenntnis*, *85*(1), 45-63. https://doi.org/10.1007/s10670-018-0017-5;

Xi, J., Si, X. A., & Malvè, M. (2023). Nasal anatomy and sniffing in respiration and olfaction of wild and domestic animals. *Frontiers in Veterinary Science*, *10*, Article 1172140. https://doi.org/10.3389/fvets.2023.1172140;

Yancy, G. (2008). Elevators, social spaces and racism: A philosophical analysis. *Philosophy & Social Criticism*, *34*(8), 843-876. https://doi.org/10.1177/0191453708094727;

Young, B. D. (2017). Enactivism's last breaths. In M. Curado & S. Gouveia (Eds.), *Contemporary perspectives in the philosophy of mind* (pp. 71-90). Cambridge: Cambridge Scholars Press. Retrieved from https://philarchive.org/rec/YOUELB;

Young, B. D. (2023). Unconsciously smelling self and others. In J. R. Searle & D. Papineau (Eds.), *Conscious and unconscious mentality* (pp. 185-206). Abingdon: Routledge. https://doi.org/10.4324/9781003409526-14;

Young, I. M. (1990). *Throwing like a girl and other essays in feminist philosophy and social theory*. Bloomington, IN: Indiana University Press.

STAVROS PANAYIOTOU
University of Athens
s.panagiotou@nup.ac.cy

A LEVINASIAN CRITIQUE OF CRITERIALISM ABOUT PERSONS[1]

abstract

In post-Kantian metaphysics, several analytic thinkers argue that criterialism is both a necessary and sufficient condition for personhood. Criterialism contends that there is a specific subset of properties that determines someone's personal identity, such as social reactive attitudes, mental complexity (second-order volition), value in combination with autonomy, self-consciousness, communication, and (practical) reason. However, Emmanuel Levinas, a prominent continental thinker, argues that behavioral properties are not tests for, but are parts of the ethical ideal of personhood. Levinas insists that the acceptance of others (as persons) cannot be a conclusion inferred from a test procedure; acceptance of others means precisely refusing to submit them to such tests. Personhood, therefore, depends on otherness. It is the other that shapes one's identity as a person, and the infinite responsibility one has towards the other person calls for an ethical shift away from the subject and the categories of being.

keywords

criterialism, Levinas, ethics, ontology, responsibility

1 This article was presented at the conference *Exploring Personal Identity: Philosophical Perspectives and Insights from the Arts*, held at Vita-Salute San Raffaele University, Milan, from October 2nd to 4th, 2024, in Section 2: *Metaphysics and Philosophy of Mind*, under topic (v): "Is it enough and/or necessary to have an articulated mental life (self-awareness, intelligence, morality, sense of responsibility, etc.) to be a person?" I would like to thank Daniel Tore for his careful proofreading and editorial assistance on the final draft of this manuscript.

Phenomenology and Mind, n. 29 - 2025, pp. 46-54
DOI: 10.17454/pam-2905
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

The purpose of this article is to explore whether any completion of the sentence "x is a person if..." may determine someone's personhood. To address this issue, it is essential to define and interpret several terms concerning personhood and personal identity. For instance, is there a specific subset of properties that determines someone's personhood? What are the conditions under which something that is a person at one time is identical with anything that exists at another time (Olson, 1997, p. 25)? Furthermore, there is an urgent need to evaluate whether it is sufficient to claim that these questions are justified within an epistemological as well as a metaphysical framework.[1] It is also necessary to articulate what constitutes their nature on *ethical* grounds.[2] This article aims to show that the analytic tradition of criterialism depicts a necessary, but not sufficient, perspective to justify the essential difference between persons, human individuals, and other creatures – particularly in ethical terms.

2. Criterialism and the Onto-logical Conception of Persons[3]

Criterialism, as a philosophical theory, contends that there is a specific subset of properties that determines someone's personhood. Criterialists claim that mental states are necessary and sufficient conditions for personhood. For example, M. Goodman writes:

> The list of conditions for personhood to be treated will be standard. Each of the conditions is to be viewed as a quality or attribute or characteristic possessed by all and only those beings regarded, and those beings to be regarded, as persons. The initial set of proposed conditions is as follows: (1) consciousness, (2) rationality, (3) ability to take and reciprocate a personal attitude toward another being, (4) the ability for complex

1 Epistemological question: How can we tell that x_t is the same as y_t. Metaphysical question: What makes it true that x_t is the same that y_t?

2 It is noteworthy that even some authors in the analytic tradition, such as D. Parfit (2011), shift their focus from personal identity to survival.

3 Concerning the conception of the person, see Mulhall and Swift, *Liberals and communitarians*, pp. 10-33. However, the notions of personhood, individualism, and atomism are quite different and complicated, especially between the philosophy of mind and social and political philosophy. Concerning the definition of atomism, see Taylor, *Philosophy and the human sciences*, pp. 187-210. Specifically, Taylor (1985) contends: "The term atomism is used loosely to characterize the doctrines of social contract theory which arose in the 17th century and also successor doctrines which may not have made use of the notion of social contract, but which inherited a vision of society as in some sense constituted by individuals for the fulfillment of ends which were primarily individual" (p. 187). Regarding the notion of individualism, see Strawson, *Individuals*, pp. 87-116.

> communication, (5) self-consciousness, (6) the ability for self-motivated activity, and (7) freedom of the will. The general consensus seems to be that this list comprises the overt necessary conditions for metaphysical personhood (1992, p. 75).

Several analytic thinkers strongly defend the above theory (Harris, 1985, pp. 16-17; Singer, 1993, p. 87; Tooley, 1972, p. 82; Warren, 1997, pp. 83-84) arguing the following: "The actual possession of the criterial properties is necessary and sufficient for membership of the primary moral constituency" (Chappell, 2011, p. 2, n. 4). Empowering this argument, and according to H. Frankfurt's view (1991), "The most essential difference between persons and other creatures can be found in the structure of a person's will" (p. 6). Frankfurt (ibid.), in addition to delineating the distinction between persons and other entities, asserts: "It is having second-order volitions and not second-order desires generally, that I regard as essential to being a person" (p. 10). In a similar manner, P. Strawson (1962) states that being a person is some kind of emergent social property: We are persons because we have and are the subject of certain reactive attitudes (pp. 1-7).[4] Whereas, G. Watson (1975) argues that volition cannot be the source of identification, as second-order desires seem to pertain not to volitions but to values (p. 216). Similarly, D. Lewis (1989) explains the act of valuing, maintaining that it is a mental, attitudinal state (p. 114). Additionally, he contends that if one does not want to desire a thing, one cannot value it. Thus, Lewis (ibid.) investigates not only second-order but also third-order or higher desires, thereby concluding that, "Valuing is just desiring to desire" (p. 115).

For criterialism, a human being is the only entity that possesses the capacity for reflective self-evaluation (Frankfurt, 1971, p. 7). Persons exhibit a unique ability to desire their desires – a characteristic that is absent in other creatures. According to Frankfurt, to qualify as a person, humans must possess a crucial presupposition: The capacity to form second-order desires or volitions regarding their will. Additionally, they merit moral consideration and respect by being autonomous. Therefore, Frankfurt concludes that only those capable of forming desires regarding their will can truly exercise free will and thus qualify as persons. This is the main reason entities lacking second-order volitions, such as wantons and robots, fail to meet the criteria for personhood.

Watson gives priority to a different presupposition of personhood in support of his theory, emphasizing practical reasoning, understood in a Platonic sense,[5] and the agent's values.[6] Watson (1975) points out that it is not volition itself but rather what is valuable and reasonable that distinguishes persons from other creatures (pp. 215-217). In a similar vein, Lewis attempts to elucidate the essential element of personhood by exploring the philosophical notion of value. The example he uses to illustrate is that of a lazy philosopher who desires fame yet remains inactive, contrasted with the mother who, despite having the first-order desire to drown her crying child, refrains from acting on it because she recognizes that it conflicts with her values. Through these examples, Lewis argues that persons shape their lives based on reasonable values, as advocated by Watson, while also considering their volitions, as proposed by Frankfurt.[7]

On the other hand, there is another branch of criterialism which postulates that, in order to identify the conditions for personhood, one must first elucidate several terms related

4 See also Hacker, *Strawson's concept of a person*, pp. 21-29.

5 Concerning Plato's account of persons see Gerson, *Knowing persons*, pp. 50-98.

6 For a detailed definition of practical reasoning see Rice, *Practical reasoning as reasoning*, pp. 49-52.

7 Concerning the relation between reason, value, and action comprising a person's will see Raz, *Engaging*, pp. 67-89.

to personal identity. Identity is commonly understood in terms of a relationship. It is the relationship that each entity bears to itself. D. Parfit asserts that identity is not a matter of degree. He contends that, while psychological continuity is crucial for personal identity, it cannot serve as its criterion because it is not a one-to-one relation. Consequently, he suggests abandoning the language of identity in favor of focusing on "survival", thereby circumventing issues with transitivity.[8] However, R. Swinburne approaches the problem of personal identity from a different perspective. He posits that personal identity is best understood in a logical sense, wherein minds or bodies – psychological and bodily continuity – may serve as indicators but are not necessarily sufficient or even necessary features for personal identity (Swinburne and Shoemaker, 1984, pp. 1-66). Swinburne believes that, contrary to Parfit, a person resembles something akin to a Cartesian soul or ego.[9] He unequivocally rejects the empiricists' conception of personal identity.[10] M. Johnston (1987) argues that personal identity is identical to personhood only if we consider persons as human biological organisms, not just psychological entities with memories and psychological traits (pp. 75-83).

3. Levinasian Ethics and the Critique of Criterialism

In post-Hegelian metaphysics, several thinkers argue that criterialism is a necessary, but insufficient presupposition for personhood. This is because the criterialist line of thought fails to provide a well-argued response to the question of whether infants, the senile, those in a persistent vegetative state, and non-human beings are to be included within the category of personhood. After all, these individuals do not meet the checklist requirements to be counted as persons, and yet we have strong intuitions that they should be. For instance, T. Chappell, a Levinasian reader, writes:

> To treat someone as a person is not to put a tick in the box by her name, to show that she has passed some inspection or met some standard of rationality or self-awareness or emotionality or whatever. Behavioral properties like rationality, self-awareness, and emotionality are not tests for, but are parts of the ideal of, personhood. To treat someone as a person is to engage with her as the kind of creature to which that ideal applies. Hence, to treat her is not, at the deepest level, a response to her behaviour at all but to her nature (...) Acceptance of others (as persons) cannot be a conclusion that we infer from a test procedure; acceptance of others means precisely refusing to submit them to such tests (2011, p. 12; 2013, p. 54).

E. Levinas (1990) argues that to consider a creature as a person involves taking an ethical attitude towards that creature, which, even before any behavioral evidence comes in, is already different from our attitudes towards creatures that (we think) are not persons (p. 295). Furthermore, the other person, that we recognize, has priority over our own personal identity, as it is through the infinite responsibility towards the other person that one finds their personhood (Capili, 2006, p. 700). Levinas (1969) vehemently rejects criterialism, as a flawed

8 For Parfit, psychological continuity is a criterion of personal identity if psychological continuity is sufficient of person identity: X is a criterion of y if x is sufficient of y. Parfit spends much ink on this issue. See for instance Parfit *Reasons and persons*, pp. 245-272; *On what matters*, pp. 83-129; *Personal identity*, pp. 3-27; *We are not human beings*, pp. 5-28; *Divided minds and the nature of persons*, pp. 91-98.

9 For instance, in Descartes, the self is the *I* of the *cogito* (*cogito ergo sum* [I think therefore I am]), the center of consciousness leading to self-awareness and intentionality. Descartes's *cogito* has been considered as the culmination of the cognitive dominance over metaphysics, aesthetics, and ethics (Casey, 2003, p. 390; Smith, 2016, p. 97).

10 Swinburne dismisses the possibility of synthesized animals or the transplantation of souls into pre-programmed machines (Swinburne and Showmaker, 1984, pp. 196-197).

theory, since it falls into the realm of immanence (p. 254). According to Levinas, human beings are not worth as much as persons' relations to each other. Infinite responsibility transcends being and its immanence (Perpich. 2008, pp. 1-16). Levinas writes:

> Responsibility goes beyond being. In sincerity, in frankness, in the veracity of this saying, in the uncoveredness of suffering, being is altered. But this saying remains, in its activity, a passivity, more passive than all passivity, for it is a sacrifice without reserve, without holding back, and in this non-voluntary – the sacrifice of a hostage designated who has not chosen himself to be hostage, but possibly elected by the Good, in an involuntary election not assumed by the elected one (1991, p. 15).

Even an infant or an individual in a persistent vegetative state can be considered a person if they are accepted as equal interlocutors (Levinas, 1998a, pp. 4-9). The other calls me to an infinite responsibility before the totality of presence through an asymmetrical and nonreciprocal relation transcending the limits of knowledge and the immanent categories of being (Bonnett, 2009). Specifically, Levinas (1985), by saying "I find myself responsible", means "[a guiltiness that is] not owing to such and such a guilt which is really mine, or to offenses that I would have committed; but because I am responsible for a total responsibility, which answers for all the others and for all in the others, even for their responsibility. The *I* always has one responsibility more than all the others" (p. 99, emphasis added). Infinite responsibility is thus by far a combination between sacrifice and holiness – defining personhood as an expression of contempt.[11] Levinas writes:

> Ethics, concern for the being of the other-than-oneself, non-indifference toward the death of the other, and hence, the possibility of dying for the other—a chance for holiness—would be the expansion of that ontological contraction that is expressed by the verb to be, dis-inter-estedness breaking the obstinacy of being, opening the order of the human, of grace, and of sacrifice [...] It is in the personal relationship, from me to the other, that the ethical 'event', charity and mercy, generosity and obedience, lead beyond or rise above being (1998a, p. 202).

According to Levinas (1998a), humans are not merely beings capable of knowing and acting through mental states, nor do they consist only of brains and psychological attitudes (p. 125). What truly makes them persons is their readiness to be sacrificed for the other person, based on an ethical, asymmetrical and non-thematizable quality (ibid., p. 168; Barber, 2020, p. 36). It is the infinite responsibility towards the other person, not a specific subset of properties, that determines someone's personhood, as "The subject is infinitely responsible for the other [...] The Levinasian subject is responsible for processes that go beyond the limits of its foresight and intention, beyond what it willed or can steer" (Yael, 2014, p. 76).

As a phenomenologist whose major interest dealt with the infinite ethical responsibility of the human subject, Levinas places major emphasis on the ethical "voice of the other" which precedes reason, and as Kleinberg-Levin (2008) correctly states: "The voices that, in their absolute singularity, come before the voice of reason, in the order of time and in the order of the ethical relation, must nevertheless, in their turn, come before the court of reason" (p. 150).

11 This affinity between Hegel and Levinas is of immense importance. Hegel's notion of personhood—as an expression of contempt—is closely related to Levinas' ethical metaphysics of *traumatic* vulnerability (Hegel, 1977, §480, emphasis added).

Above and beyond fundamental ontology, Levinas states that ethics precedes any ontological *a priori* subset of mental properties (Erdur, 2020, pp. 273-276). He also emphasizes that the individual's acceptance of responsibility in engaging with the other arises from recognizing the alterity of the other as taking precedence over the rights of the self (Greisch and Rolland, 1993). On this note, Levinas (1994) defines ethical responsibility as follows: "To hear a voice speaking to you is *ipso facto* to accept obligation toward the one speaking" (p. 48).

Levinas attempts to transcend the notion of personhood on ethical grounds, refusing to reduce intuition to immanence. He avoids relying on ontological criteria such as consciousness and rationality, instead proposing ethical criteria to justify primary moral constituency. He argues that the priority of infinite responsibility for the other person is both a necessary and sufficient condition for personhood (Zhao, 2020, pp. 257-261). In contrast to criterialists' features of mental states, such as communication and action, as well as their insistence on ordinary morality, such as moral consideration and respect, Levinas "Speaks of a responsibility that goes beyond what it is possible to do, beyond my actions and their consequences, beyond the distinction between the voluntary and the involuntary, to an infinite and irrecusable responsibility for the other" (Perpich, 2008, p. 83). The other always exists outside the domain of the "same" and, therefore, outside totality. Ethically, this means that no rational system can supersede the call to infinite responsibility that emerges from encountering the "face," which addresses the individual from a completely external position. The call of the "face" is profoundly objective and beyond the self's ability to understand and relativize the external dimension of being (Bloechl, 2000).

We can thus infer that it is not sufficient to merely have a specific subset of ontological properties based on knowledge and consciousness to be a person. Levinas (1998a) raises several objections to this argument, stating: "Knowledge only comprehends itself in its own essence, starting from consciousness, whose specificity eludes us when we define it with the aid of the concept of knowledge which itself supposes consciousness" (p. 58). Therefore, what makes me a person is the Levinasian ethical potentiality of a nonreciprocal asymmetrical relationship with the other and the infinite responsibility thus cast upon *me*. As M. Barber writes:

> The other is experienced as asymmetrically inviting me to be responsible, and there are various modes in which that responsibility can be fulfilled, only one of which is to enter into speech or reasoning with the other. Thus, my relationship with the other establishes a saying relationship, which affords the context within which anything is said and which, therefore, forms the setting for philosophy, science, and the processes of justification. For Levinas, insofar as such processes take place on a terrain common to me and the others, they indicate that the asymmetrical relationship of proximity has already been modified by the appearance of the equalizing Third person (2008, p. 632).

Levinas gives a major priority to relational intersubjectivity over fundamental ontology in order to define what it means to be a person. As M. Ruti (2015) correctly states: "The properly human, for Levinas, therefore begins when I transcend my ontology and begin to exist for the other. In this sense, ethics brings about a kind of crisis, upheaval, or suspension of being, which jolts me out of my absorption in my own life" (p. 6). Therefore, for Levinas, relationalism precedes individualism in all facets of the philosophy of personhood. And finally, as Chappell (2020) underlines: "If relationalism about persons is true, then criterialism about persons looks not just false but incoherent: It takes as a criterion of personhood what it only makes sense to treat as an ideal of personhood (p. 188)".

4. Conclusion

In summation, we can conclude that for Levinas, criterialism is fundamentally flawed because it fails to consider the "otherness of the other" as a criterion of personhood. Therefore, it is not sufficient to have an articulated mental life, that is, self-awareness, intelligence, morality, and second-order volition, to be a person. Levinas aims to show that relationalism precedes individualism. Everything starts from otherness and not from being itself. The experience and the perception of otherness, which cannot be absorbed by mental states, is, in Levinas's view, the basic idea of the philosophy of personhood. Levinas assigns primacy to external alterity, which is not a feature of knowledge but of ethics. In fact, radical alterity is the alterity of the absolute relation to the other, which demands our ethical response rather than a notion of "pure Being" as some impersonal inner reality. The thinking subject enters into a relationship with human beings escaping the representation and totalization of ontology. For Levinas, the thinking subject calls itself to an infinite responsibility beyond itself, which exists prior to mental states. Levinas is thus convinced that it is not sufficient to possess mental states in order to be considered as persons; it is the other who gives us this privilege.

REFERENCES

Barber, M. (2020). The second source of normativity and its implications for reflective endorsement: Levinas and Korsgaard. In M. Fagenblat & M. Erdur (Eds.), *Levinas and analytic philosophy* (pp. 29-54). New York, NY: Routledge. doi:10.4324/9780429462580;

Barber, M. (2008). Autonomy, reciprocity and responsibility: Darwall and Levinas on the second person. *International Journal of Philosophical Studies*, 16(5), 629-644. doi:10.1080/09672550802493728;

Baxter, R. (1988). Identity in the loose and popular sense. *Mind*, 97, 575-582. doi:10.1093/mind/XCVII.388.575;

Bloechl, J. (2000). *The face of the other and the trace of God*. New York, NY: Fordham University Press;

Bonnett, M. (2009). Education and selfhood: A phenomenological investigation. *Journal of Philosophy of Education*, 43(3), 357-370. doi:10.1111/j.1467-9752.2009.00698.x;

Capili, A. (2006). Emmanuel Levinas and the human person: On the philosophical conditions of war and peace. *Revista Portuguesa de Filosofia*, 62(2), 697-711;

Casey, T. (2003). Levinas' idea of the infinite and the priority of the other. *Gregoriarum*, 84(2), 383-417;

Chappell, T. S-G. (2013). Knowledge of persons. *European Journal for Philosophy of Religion*, 5(4), 31-56. doi:10.24204/ejpr.v5i4.203;

Chappell, T. S-G. (2011). On the very idea of criteria of personhood. *The Southern Journal of Philosophy*, 49(1), 1-27. doi:10.1111/j.2041-6962.2010.00042.x;

Chappell, T. S-G. (2020). Personal knowledge. In M. Fagenblat & M. Erdur (Eds.), *Levinas and analytic philosophy* (pp. 166-189). New York, NY: Routledge;

Erdur, M. (2020). Levinasian ethics as first philosophy. In M. Fagenblat & M. Erdur (Eds.), *Levinas and analytic philosophy* (pp. 269-278). New York, NY: Routledge;

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5-20. doi:10.2307/2024717;

Gerson, L. (2003). *Knowing persons: A study in Plato*. Oxford: Oxford University Press;

Goodman, M. (1992). A sufficient condition for personhood. *The Personalist Forum*, 8(1), 75-81;

Greisch, J., & Rolland, J. (1993). *Emmanuel Lévinas: L'éthique comme philosophie première*. Paris: Éditions du Cerf;

Hacker, P. (2002). Strawson's concept of a person. *Proceedings of the Aristotelian Society*, 102(1), 21-40;

Harris, J. (1985). *The value of life: An introduction to medical ethics*. London: Routledge & Kegan Paul. doi:10.4324/9780203005606;
Hegel, G.W.F. (1977). *The phenomenology of spirit*. Trans. A. Miller. Oxford: Oxford University Press;
Johnston, M. (1987). Human beings. *The Journal of Philosophy*, 84(2), 59-83;
Kleinberg-Levin, D. (2008). *Before the voice of reason*. New York, NY: SUNY Press;
Levinas, E. (1969). *Totality and infinity*. Trans. A. Lingis. Pittsburgh, PA: Duquesne University Press;
Levinas, E. (1985). *Ethics and infinity: Conversations with Philippe Nemo*. Trans. R. Cohen. Pittsburgh, PA: Duquesne University Press;
Levinas, E. (1990). *Difficult freedom: Essays on Judaism*. Trans. S. Hand. Baltimore: The Athlone Press & John Hopkins University Press;
Levinas, E. (1991). *Otherwise than being, or, beyond essence*. Trans. A. Lingis. Dordrecht: Springer;
Levinas, E. (1994). *Nine talmudic readings*. Trans. A. Aronowicz. Bloomington, IN: Indiana University Press;
Levinas, E. (1998a). *Entre nous: On thinking-of-the-other*. Trans. M. B. Smith & B. Harshav. New York, NY: Columbia University Press;
Levinas, E. (1998b). *Of God who comes to mind*. Trans. B. Bergo. Stanford, CA: Stanford University Press. doi:10.1515/9781503616882;
Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society*, Suppl. Vol., 113-138. doi:10.1093/aristoteliansupp/63.1.89;
Mulhall, S., & Swift, A. (1992). *Liberals and communitarians*. Oxford: Blackwell;
Nielsen, L. (2023). *Bundle theory and loose identity*. PhD dissertation. Olomouc: Palacky University;
Noonan, H. (2019). Personal identity: The simple and complex views revisited. *Disputatio*, 11(52), 9-22. doi:10.2478/disp-2019-0001;
Olson, E. (1997) *The human animal: Personal identity without psychology*. New York, NY: Oxford University Press. doi:10.1093/0195134230.001.0001;
Parfit, D. (1971). Personal identity. *The Philosophical Review*, 80(1), 3-27. doi:10.2307/2184309;
Parfit, D. (1984) *Reasons and persons*. Oxford: Clarendon Press;
Parfit, D. (2011). *On what matters*. New York, NY: Oxford University Press. doi:10.1017/S0031819112000423;
Parfit, D. (2012). We are not human beings. *Philosophy*, 87(1), 5-28. doi:10.1017/S0031819111000520;
Parfit, D. (2016). Divided minds and the nature of persons. In S. Schneider (Ed.), *Science fiction and philosophy: From time travel to superintelligence* (pp. 91-98). New York, NY: Wiley. doi:10.1002/9781118922590.ch8;
Perpich, D. (2008). *The Ethics of Emmanuel Levinas*. Stanford, CA: Stanford University Press. doi:10.1515/9780804779784;
Raz, J. (2002). *Engaging reason: On the theory of value and action*. New York, NY: Oxford University Press. doi:10.1093/0199248001.001.0001;
Rice, H. (1989). Practical reasoning as reasoning. *Proceedings of the Aristotelian Society*, 89(1), 49-63. doi:10.1093/aristotelian/89.1.49;
Ruti, M. (2015). *Between Levinas and Lacan: Self, other, ethics*. New York, NY: Bloomsbury;
Schonfeld, E. (2007). *The wonder of subjectivity: A reading of Levinas's philosophy*. Tel Aviv: Resling;
Singer, P. (1993). *Practical ethics*. Cambridge: Cambridge University Press;
Smith, F. (2016). Gadamer and Levinas on the topic of sociality. *Sophia Philosophical Review*, 9(2), 95-109;
Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 1-25;

Strawson, P. (2006). *Individuals*. New York, NY: Routledge;
Swinburne, R., & Showmaker, S. (1984). *Personal identity*. Oxford: Blackwell. doi:10.1017/S0028428900033151;
Taylor, C. (1985). *Philosophy and the human sciences*. Oxford: Oxford University Press;
Tooley, M. (1986). Abortion and infanticide. In P. Singer (Ed.), *Applied Ethics* (pp. 57-86). Oxford: Oxford University Press. doi:10.1515/9780691233161-004;
Vezer, M. (2007). On the concept of personhood: A comparative analysis of three accounts. *Lyceum*, 9(1), 12-27. doi:10.5840/lyceum2007913;
Watson, G. (1975). Free agency. *Journal of Philosophy*, 72, 205-220. doi:10.2307/2024703;
Warren, M. A. (1997). On the moral and legal status of abortion. In H. Lafollette (Ed.), *Ethics in practice: An anthology* (pp. 91-102). Oxford: Blackwell. doi:10.5840/monist197357133;
Yael, L. (2014). *The intersubjectivity of time: Levinas and infinite responsibility*. Pittsburgh, PA: Duquesne University Press;
Zhao, G. (2020). *Between virtue and the theory of subjectivity*. In M. Fagenblat & M. Erdur (Eds.), *Levinas and analytic philosophy* (pp. 253-268). New York, NY: Routledge.

SECTION

SECTION 2

METAPHYSICAL QUESTIONS ABOUT MIND, BRAIN, AND SELF

ANDREA C. M. BOTTANI
University of Bergamo
andrea.bottani@unibg.it

ON THE SUBSTANTIVE IRRELEVANCE OF REDUCTIONISM ABOUT PERSONS[1]

abstract

The contrast between reductionist and anti-reductionist views of persons (respectively also known as "complex" and "simple" views) has shaped the metaphysical debate about personal identity over the last decades, the idea being that no difference between views of persons can be more crucial and important than their reductionist (or anti-reductionist) attitude. I want to argue that the substantive importance of reductionism about persons has been overestimated. However important the contrast between simple and complex views of persons may be on a meta-metaphysical level, it has surprisingly little bearing on the substantive issue of the nature of persons and of their identity conditions across time.

keywords

person, metaphysics, reductionism, materialism, persistence

1 Thanks to Richard Davies and an anonymous referee for useful comments on earlier versions of this paper.

Phenomenology and Mind, n. 29 - 2025, pp. 58-66
DOI: 10.17454/pam-2906
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

I. Persons, reductionism and empty questions

Very often, philosophers ask questions of the form 'What is X?' – for example, 'what is the good?', 'What is justice?', 'What is truth?'. There is nothing surprising about that, for philosophy is, in a sense, just the rational effort to clarify the deep nature of things. But questions of the form 'What does "X" mean?' – for example, 'what does "good" mean?' 'what does "justice" mean?', 'what does "truth" mean?' – are also common in philosophy. And one might wonder whether, in what cases and to what extent questions of the former kind can be reduced to questions of the latter. This is an important meta-philosophical issue, about which I shall say little if anything here. Let me just observe that the distinction between questions about nature and questions about meaning appears clear-cut and well founded in a large number of contexts. It would be crazy, for example, to try to discover the nature of pain by reading in a dictionary what 'pain' means. In other cases, however, the distinction appears less clear. A famous case concerns ships. Plutarch writes:

> 'The ship on which Theseus sailed with the youths and returned in safety, the thirty-oared galley, was preserved by the Athenians down to the time of Demetrius Phalereus. They took away the old timbers from time to time, and put new and sound ones in their places, so that the vessel became a standing illustration for the philosophers in the mooted question of growth, some declaring that it remained the same, others that it was not the same vessel' (Plutarch, 1914, *23, 1*).

In more recent centuries, following an ancient suggestion, Thomas Hobbes imagined in his *De Corpore* that the ship's old planks had been collected on land, and reassembled just as they were arranged in the original ship (Hobbes, 1999, 2, 11, 7). If this were to happen, he asked, which ship would be the original one? The one preserved by the Athenians or the reassembled one? Although Hobbes argued that the reassembled ship is the true original,[1] this may depend more on language than on reality. Consider that you can know everything there is to know about each particular wooden plank of the ships in question and still have no idea of which ship is the original one. So one might say: look, it just depends on what you mean by

1 See also Hobbes, 1976: 'If some part of the first material has been removed or another part has been added, that ship will be another being, or another body. For, there cannot be a body "the same in number" whose parts are not all the same, because all a body's parts, taken collectively, are the same as the whole (12,3).

"ship"! If you conceive of a ship as something invariable in its material composition, then the reassembled ship is the true original. If you conceive of a ship as something spatio-temporally continuous, then the original ship is the preserved one.

It has been suggested that the reason why questions about nature and questions about meaning tend to merge in the case of ships is that ships are *reducible to other things*. They ultimately consist in things that are not ships (wooden planks and the like). Some philosophers are as reductionist about persons as they are about ships. They believe that, ultimately, persons are nothing over and above things that are not persons.[2] In their view, a complete catalogue of what there is in heaven and earth can fail to mention persons, and a complete description of reality can be formulated in impersonal terms. And they think that questions about the nature of reducible entities are more about meaning than about reality.

The very idea of a reduction is widely discussed and highly controversial, and the relations among reduction, emergence, supervenience, ontological dependence and so on are far from clear. This is an intricate and slippery matter, and it is not among my aims to take discussion of such kind of issues into consideration. Let me just assume that reducing things of kind F to things of kind G is not claiming that the Fs do not exist, so that reductionism about the Fs is not eliminativism about them. Thus, a one-category ontology, which reduces everything that exists to things of just one category (say, particulars, universals, or tropes) need to be revisionary[3] – i.e., to deny the existence of a wide number of commonsensical categories of objects – but aims only to explain their nature in purportedly more basic and fundamental terms. On the one hand, reduction can work as a tool for explanation, taxonomical simplification and ontological deflation, which explains why reductionism is a very natural ontological attitude so long as ontological economy is seen as a theoretical *desideratum*. To reduce things of one category to things of another helps to decrease the global number of categories, enabling us to improve what David Lewis calls "qualitative parsimony". On the other hand, reductionism mitigates the impact of metaphysical analysis on our ordinary ontological commitments, helping to reconcile the metaphysical image of the world with what Sellars called "the manifest image" of it.

The contrast between reductionist and anti-reductionist views of persons has shaped the discussion about the nature of persons over the last decades. Starting from Parfit (1984), the contrast between so-called "complex" (or reductionist) and so-called "simple" (or anti-reductionist) views of persons has been taken as a fundamental divide between two kinds of deeply different conceptions of persons: on the one hand, the reductionist (or complex) views, according to which persons are nothing over and above things that are not persons and facts about personal identity across time are nothing over and above impersonal facts; on the other hand, the anti-reductionist (or simple) view, according to which persons consist in nothing but themselves and there is nothing impersonal in which facts about personal identity across time can consist. In this view, the alternative between simple and complex theories of persons is an

2 It is natural to distinguish between two questions. 1) What necessary and sufficient conditions must be satisfied for a person existing at a time t^1 and a person existing at a subsequent time t^2 to be the same person? 2) What necessary and sufficient conditions must be satisfied for something whatever to be a person? Philosophers working in the analytic tradition have often obscured the distinction between these questions, treating the latter nearly as the shade of the former. Reductionism about persons has been accordingly interpreted as the thesis that personal identity across time consists in certain relations between impersonal entities. Although the correctness of this approach is certainly disputable, nothing substantial hinges here on accepting or refusing it (for a discussion, see Tomasetta, 2015).

3 On the distinction between descriptive and revisionary metaphysics, see Strawson, 1959.

extremely important one on which much depends, and any other difference between theories of persons, however important, is less basic and fundamental.[4]

I want to argue that the substantive importance of this contrast has been overestimated. Philosophical conceptions of persons can differ in a number of respects that are more important and fundamental than their attitude to reducing (or resistance to reducing) persons to purportedly more basic entities. Philosophers give different answers to substantive questions such as: are persons material or not? Do they belong to the natural domain or not? Are they indeterminate or not? Do they endure, perdure or exdure? But, as I shall argue, neither reductionism nor anti-reductionism about persons can dictate on their own any precise answers to these questions. However important the contrast between simple and complex views of persons may be on a meta-metaphysical level, it is on its own of surprisingly little substantive import. In particular, there is no logical or semantical tie between reductionism about persons as such and the idea that personal identity across time is indeterminate (or even that it is indeterminate whether some entity is a person or not). Likewise, reductionism about persons entails neither materialism nor naturalism about ourselves (both in a weak and in a strong sense of 'materialism' and 'naturalism'). And, conversely, neither materialism nor naturalism about ourselves entails a reductionist view of our nature. In addition, in no way does reductionism about persons entail that we are dispensable and unimportant, or that, so to say, we should not be taken seriously, ontologically speaking. Finally, neither reductionism about persons nor its denial can dictate how persons persist across time (i.e., whether they perdure, endure or exdure).

A paradigmatic reductionist approach to personal identity was offered in the 1980s by Derek Parfit, in his famous book *Reasons and Persons*.[5] In this book, Parfit defends reductionism about persons and their identity, which he calls "the Complex View", against anti-reductionism about persons and their identity, which he calls "the simple view". He formulates the complex view as the thesis that facts about personal identity consist in other impersonal facts. And he argues, more specifically, that facts about personal identity are nothing over and above facts about psychological connectedness and continuity. As is well known, the core of his psychological reductionism is a certain inversion of the dependence relation between a subject and her mental states. The existence of a mental state is no longer said to require the existence of a subject, it is rather the existence of a subject that is said to require the existence of a number of adequately related mental states. Roughly, the idea goes as follows. First, mental states existing at the same time are unified as a maximal group of simultaneous mental states by their co-consciousness (where the notion of co-consciousness of two or more mental states must be assumed as primitive and intuitively clear, and not defined as consciousness of the same subject, to avoid circularity). Second, two maximal groups of simultaneous co-conscious mental states existing at different times are *of* the same subject just in case a sufficient number of mental states of the former group are causes or concauses of a sufficient number of mental states of the latter group - and, in addition, there is enough qualitative similarity between mental states of the former group and mental states of the latter. Since no precise threshold is given beyond which qualitative similarity-*cum*-causal connectedness of two groups of mental states is "sufficient" in order for them to be *of* the same person, personal identity across time can only be indeterminate. But this indeterminacy

4 The alternative is often seen as an indispensable key for grasping the ultimate meaning of the problem of personal identity, and of the discussion about it. A good example is Gasser and Stefan's reading on personal identity (Gasser and Stefan, 2012).

5 Some of the main ideas of the book were partially developed by Parfit in the seventies in some papers, among which Parfit, 1971.

has nothing ontic about it, it is just the indeterminacy of a certain way of describing some phenomena – in personal rather than impersonal terms. Ultimately, an 'empty question' is just one whose answer depends on semantics rather than on reality.

II. Complex yet immaterial determinate persons

No doubt, Parfit's reductionism treat persons as no less indeterminate, strictly material and ontologically dispensable than ships, but it is not *qua* reductionist that it does so. Let me begin by showing why reductionism about persons does not on its own imply that persons are indeterminate and, conversely, antireductionism about persons does not on its own imply that persons are determinate.

Suppose one supplements the meaning of "psychologically connected" with a precise criterion for establishing to what degree two persons are psychologically connected, and Parfit's criterion of personal identity with a precise specification of the minimum degree to which two persons must be psychologically connected in order to be the same person. This supplemented quasi-Parfitian view would still be complex, but would no longer entail that persons are indeterminate. Conversely, imagine one believes at the same time that there is something as ontic vagueness, that persons are simple immaterial beings and that simple immaterial beings are vague entities. Or else, suppose one believes that persons are some sorts of fundamental, ontologically indeterminate quantum objects. In both cases one would believe that persons are simple and indeterminate.

Likewise, reductionism about persons as such does not imply materialism about them and, conversely, anti-reductionism about persons as such does not imply anti-materialism about them. To begin with, consider a Berkeleyan version of the psychological criterion of personal identity conceiving of mental states and events as purely immaterial ideas and thoughts rather than neurophysiological states and events. This unorthodox version of the psychological criterion would plainly be no less reductionist than the standard familiar version, but it would be radically anti-materialist and anti-naturalist as much as the standard version is radically materialist and naturalist. It is far from clear that this imaginary anti-materialist version of the psychological criterion of personal identity would be paradoxical or incoherent, even though its coherence with contemporary sciences may of course be put in doubt.

Again, imagine a theory according to which (i) persons are complexes of (say) four very thin immaterial atoms, (ii) they have the same components at any moment of their life, and (iii) the identity over time of their atomic components is brute and non-analysable (suppose the four atomic parts are conceived of as something like self-subsistent pure forms, souls or the like). Although, as far as I know, nobody has ever embraced (i)-(iii), a theory incorporating these theses would entail that persons are nothing over and above things that are not persons, and their identity across time consists in the diachronic identity of things that are not persons. But still, that reductionist theory would treat persons as purely immaterial entities.

Examples of this sort clarify a point that should be evident *a priori*. The idea that persons are nothing over and above things that are not persons need not be either materialist or anti-materialist *per se*, nor either naturalist or anti-naturalist *per se*. This just depends on the nature of the non-personal entities to which persons are purportedly reduced. Since, in principle, they can be either material or immaterial, they can be members of the orthodox ontology of natural sciences or not, and they can belong, in general, to a wide number of heterogeneous domains, it is completely unmotivated to think that reductionism can be provided on its own of any substantive value.

III. Irreducible yet material persons

The case of a materialist anti-reductionist theory of persons is less exotic. To make just a couple of examples, both Lynne Rudder Baker and Jonathan Lowe, have proposed in recent years approaches to the nature of persons that are anti-reductionist but mildly materialist.

The same may also be true, at least in a possible version, of some narrativist approaches to diachronic personal identity.

According to Baker, I bear to my living body the same relation that a bronze statue bear to the matter of which it is made.[6] A human animal constitutes me, but I am not identical with it, just as the statue is not identical with the bronze, even if there is a sense of 'is' (the so-called 'is' of composition) in which the statue can correctly be said *to be identical* with the bronze. The statue is not numerically identical with the bronze because it has essential properties that the bronze lacks – in particular, the shape that the sculptor has given to the bronze, making the statue just the statue it is. Likewise, I have essential properties that my body lacks – according to Baker, in particular, my ability to refer to myself as "I", to perceive the experiences and perceptions that I happen to have as *mine*, to refer to any other people as "you", and so on. In short, what Baker calls "first-person perspective". In Baker's view, a person existing at t^1 is identical with a person existing at t^2 just in case they have the same first-person perspective, but there is nothing more basic in which having the same first-person perspective consists. This makes her approach at least similar to what Parfit calls "the simple view", for in Baker's view there is ultimately nothing in which facts about personal continuity across time can consist. So, leaving aside its possible difficulties, Baker's approach is anti-reductionist, but it is perfectly compatible with materialism, for it has no need to appeal to immaterial souls and assumes that we are essentially embodied, i.e. necessarily constituted by something material. However, it is anti-naturalistic in orientation.[7] In fact, as Baker herself remarks the very idea of a first-person perspective is an impossible challenge to naturalism, for naturalistic theories are rentlessly third-personal in character.

A different version of a materialist but non-reductionist approach to the nature of persons, sometimes called 'Brobdingnagian atomism', was also defended by Jonathan Lowe since the nineties in a number of works.[8] According to Lowe, I am a material entity located in the extended region in which my body is also located, but I am not identical with my body, for my body has parts while I do not. The idea of an extended atom, indivisible even though located in a divisible spatial region and locally differentiated in the various parts of the occupied region is extremely difficult to be grasped. How can I have a colour where there are my eyes and another colour where there is my neck if not by having different parts of different colours? And how can I weigh 86 kgms if not by having exactly 86 proper disjoint parts weighing 1 kg each?

One tentative way to clarify the point is by means of a temporal analogy. Endurantists treat ordinary objects as enduring, i.e. temporally atomic even though temporally persistent. In their view, I have no temporal parts, so I am entirely present at each of the different moments at which I exist, even though I have different qualities at those moments. I may be conceived of likewise, as entirely located in every point of an extended region, having different qualities in many of those points. Just as endurantists deny that I divide across time, having different parts at different moments, Lowe denies that I divide in any of the three ordinary spatial dimensions. In short, I am a 0-dimensional brobdingnagian extended atom. However, this way to clarify Lowe's view might be flawed after all, and a Brobdingnagian atom, contrary to an enduring thing, might instead fail to be multi-located. It might uniquely exist in the entire spatiotemporal region where it extends and in no proper part of it, and inherit its local properties from the parts of the co-located hunk of matter.

6 Baker, 1997, 2000.

7 See especially Baker, 2013.

8 See in particular Lowe, 2001.

Be that as it may, Lowe offers an argument to the effect that we are Brobdingnagian atoms. The argument appears formally correct, even though its premises are extremely controversial. Let me skip this. What matters in this context is that Lowe's Brobdingnagian atomism is an anti-reductionist but materialist account of our own nature. In this respect, Lowe's approach converges with Baker's, but diverges from it inasmuch as it denies that I have parts. Also, Lowe's approach is not so clearly anti-naturalistic as Baker's, for both multi-located entities and extended simples may be not so clearly excluded by our best scientific theories (in particular by quantum-mechanics), however unpalatable they may be at the macro-level.

One might ask: is this kind of mildly materialist anti-reductionism about persons also a fundamental aspect of hylomorphic conceptions of our own nature - i.e., the general idea that the relation between a person and her body is a particular case of the matter-form relation?[9] One might tentatively answer: only inasmuch as forms are not treated as ontologically independent from matter in the case of persons, i.e., only inasmuch as they are not treated as self-subsistent. If, on the contrary, forms come to be seen as some sort of self-subsistent souls capable of disembodied existence, as in Aquinas' view, in no way can hylomorphism be a sort of mild materialism. As is well known, however, hylomorphism is as variable as the very notion of form, and can accordingly admit a great number of substantially divergent versions.

On the other hand, the inclusion of narrative ties in the domain of personal continuity relations across time (see for example Slors, 1998), is *prima facie* likely to weaken the hard reductionism implicit in psychological criteria. Since narration develops through intentional and semantic rather than merely causal ties, it can in principle be interpreted as narration *of a subject*, who has therefore to be imagined as already given *ab initio*. But ties of that sort might also be conceived of as naturalizable in a third-person perspective, as in Dennett's idea of the self as the centre of narrative identity.

IV. Can persons matter in any way if they consist in something impersonal?

Even granted all this, one may nonetheless insist that there is at least one substantive consequence about ourselves that reductionism about persons must have as such. If persons are nothing over and above things that are not persons, a complete inventory of what exists in heaven and earth will not include me, and a complete description of reality can fail to mention me. If so, however, I am so irrelevant and unimportant as to be absent from the layer of what really matters in the world. But this may be false, for the catalogue can include me even if persons are nothing over and above things that are not persons.

Consider in this connection Peter van Inwagen's idea that necessarily, if x is a person existing at time t^1 and y exists at time t^2, then $x = y$ just in case x's biological life at t^1 is identical to y's biological life at t^2, where a biological life is a self-sustaining chemical process that always coincides with a living organism.[10] If this is true, I am a human animal who happens to be now (but not some decades ago) a person. This account of my nature says nothing fundamental about what it is to be a person. It says that human persons are identical to human animals but does not say that being a human person consists in being a human animal. If being persons requires being conscious, having desires, aims and so on, there are human animals that are not persons (for example, a three month human embryo), as well as human animals that have never been, and will never be, persons (for example, a human embryo miscarried at three months). So, some human animals are not persons. And perhaps, conversely, some persons are not human animals, if there are or could be non-human or non-organic persons – e.g., disembodied minds, angelic and divine persons, artificial minds (if there are

9 See for example Stump, 1995.
10 Van Inwagen, 1990, 2007.

any). Therefore, being a human animal is neither necessary nor sufficient for being a person and, at least in principle, the animalist can completely endorse Parfit's answer to the question "In what does it consist to be a person?". So, he can be no less reductionist about persons than the friend of the psychological criterion. Since I am a human animal, however, I continue to be included in the catalogue, and a complete description of the universe cannot fail to mention *me*. This shows once again how a reductionist theory of persons can fail to have any substantive consequences about our own nature.

V. Do simple persons necessarily endure?

It may seem natural to believe that, if persons are fundamental irreducible entities, they can only persist by enduring, i.e., by being entirely present at any time at which they exist. The purported reason is that perduring entities are mereological sums of temporal parts and parts are most naturally seen as more fundamental than the whole they are parts of (after all, the existence of a whole may seem to depend on the existence of each of its parts in a way in which the existence of any of its parts does not seem to depend on the existence of the whole).

If this is true, the alternative between simple and complex views of persons profoundly affects the substantive issue of how persons persist across time. But there are doubts that this is so. On the one hand, persons might fail to endure by exduring instead of by perduring. An exduring thing is a stage, i.e., an instantaneous object existing on its own at just one time (at any other time it can only exist by proxies – instantaneous representatives existing on their own at just that time).[11] Since an object of this sort is a simple, non-reducible entity, simple persons might exdure, and so persist without enduring.

On the other hand, the idea that the parts are prior to the whole is highly controversial, insofar as friends of priority monism believe exactly the contrary: it is the whole that is prior to its parts, not the reverse (see for example Schaffer, 2010). If priority monists are right, a person might have temporal parts while being a basic, irreducible thing – and so a *simple* entity at the fundamental level. But even if priority monists were wrong and nothing irriducible could have any parts, simple persons could perhaps perdure in a non-mereological sense, in which something can perdure without having any temporal parts. Of course, if a perduring object is conceived of as a mereological sum of its temporal parts, nothing simple can perdure. But the very idea of a perduring object can be characterized in locational rather than mereological terms (i.e., in terms of the relations that persisting objects bear to the spacetime region(s) where they are located rather than in terms of part-whole relationships).

Say that an entity *x* is *exactly located* in a (spacetime) region r if and only if *x* has exactly the same shape and size as r (in other words, just in case r is *x*'s 'shadow' in spacetime). And say that *x* is *totally located* in a region r if and only if r is the fusion (i.e., the sum) of all the regions r' such that *x* is exactly located in r'. The region r in which x is totally located is often called "x's path".[12]

Given these definitions, the very concept of a perduring object can be rephrased in locative terms as follows: a persisting object *perdures* just in case its total location and its exact location are identical (i.e., just in case it is exactly located at its path and nowhere else). As can easily be seen, something can perdure in this sense even if it is an extended simple, i.e. no part of it is located in any proper part of its path. Since what perdures in this way while having no temporal parts is an extended atom, it is simple and does not consist in anything else. Therefore simple persons might perdure in this way.

11 Hawley (2001).

12 See Gilmore, 2006; Parsons, 2008.

I've argued that, contrary to the mainstream, the contrast between so-called simple and so-called complex views has little bearing on the substantive problem of our own nature. This is not to say that it has no bearing at all. For example, if one rejects priority monism and conceives of perdurance in mereological terms, the 'simple' thesis (unlike the 'complex' one) entails that persons do not perdure. If one doesn't believe in ontic vagueness, the 'simple' thesis (unlike the 'complex' one) entails that they are determinate. And so on and so forth. Therefore, the complex and the simple theses do not inferentially behave in the same way – i.e., they do not always yield the same substantive conclusions in conjunction with the same premisses. Taken by themselves, however, they say very little about our own nature. However important their contrast may be on a meta-metaphysical level, it is of surprisingly little substantive import.

REFERENCES

Baker, L.R. (1997). Why Constitution is not Identity. *The Journal of Philosophy*, 94 (12), 599-621;
Baker, L.R. (2000). *Persons and Bodies. A constitution View*. Cambridge: Cambridge University Press;
Baker, L.R. (2013). *Naturalism and the First Person Perspective*. Oxford: Oxford University Press;
Berkeley, G. (1998). *A Treatise Concerning the Principles of Human Knowledge*. Ed. J. Dancy. Oxford: Oxford University Press;
Gasser, G., & Stefan, M. (Eds.) (2012). *Personal Identity. Complex or Simple?*. Cambridge: Cambridge University Press;
Hawley, K. (2001). *How Things Persist*. Oxford: Oxford University Press;
Hobbes, T. (1976). *Thomas White 'De Mundo' Examined*. London: Bradford University Press;
Hobbes, T. (1999). *Human Nature and De Corpore Politico*. Oxford: Oxford's World's Classics;
Lowe, J. (2001). Identity, Composition and the Simplicity of the Self. In K. Corcoran (Ed.), *Soul, Body and Survival. Essays on the Metaphysics of Human Persons*. Ithaca, NY: Cornell University Press;
Parfit, D. (1971). Personal Identity. *Philosophical Review*, LXXX, 3-27;
Parfit, D. (1984). *Reasons and Persons.* Oxford: Oxford University Press;
Parsons, J. (2008). Theories of Location. In D. Zimmerman (Ed.), *Oxford Studies in Metaphysics* (pp. 201-232). Oxford: Oxford University Press;
Plutarch (1914). *Plutarch's Lives*. English Trans. B. Perrin. Cambridge, MA: Harvard University Press;
Schaffer, J. (2010). Monism: the Priority of the Whole. *Philosophical Review*, Vol. 119, No. 1, 31-76;
Slors, M. (1998). Two Conceptions of Psychological Continuity. *Philosophical Explorations*, *1*(1), 61-80;
Strawson, P. (1959). *Individuals. An Essay in Descriptive Metaphysics*. London: Methuen;
Stump, E. (1995). Non-Cartesian Substance Dualism and Materialism without Reductionism. *Faith and Philosophy*, 12 (4), 505-531;
Tomasetta, A. (2015). *Persone umane. Teorie contemporanee in metafisica analitica*. Roma: Carocci;
Van Inwagen, P. (1990). *Material Beings*. Ithaca, NY: Cornell University Press;
Van Inwagen, P. (2007). A Materialist Ontology of the Human Person. In P. van Inwagen & D. Zimmerman (Eds.), *Persons. Human and Divine*. Oxford: Oxford University Press.

ALBERTO BARBIERI
Vita-Salute San Raffaele University
barbieri.alberto@unisr.it

GROUNDING OUR SENSE OF PERSONAL EXISTENCE: HOW NOT TO DO IT[1]

abstract

In contemporary analytic philosophy of mind, the sense of our personal existence that several classical philosophers have believed to permeate our experience is typically cashed out in terms of the ubiquity of our inner awareness of our own experience. In this paper, I address the issue of what grounds such an inner awareness, arguing against the widespread view that it obtains in virtue of a more fundamental awareness the occurrent experience has of itself. This is the state self-awareness view (SSV). Specifically, I argue that all versions of SSV are bedevilled by what I call the problem of state *awareness, that is, the problem of understanding why and how a mental state-level relation like state self-awareness could recover a subject-level relation like inner awareness. I also defend the depth of this problem by blocking a potential solution.*

keywords

sense of existence, inner awareness, for-me-ness, consciousness, state self-awareness view, self-representationalism

1 This paper was presented at the San Raffaele School of Philosophy 2024. I am grateful to all the participants for their engagement and observations. For thorough comments on a previous draft, I am especially grateful to Elisabetta Sacchi. I would also like to thank an anonymous referee of *Phenomenology and Mind* for their useful comments and remarks. This work has been supported by the PRIN Project 2022YXBJWX "Higher-order Perceivable Properties" (HOPP), funded by the European Union—Next Generation EU, Mission 4 Component 2 Investment 1.1, CUP D53D23009960008.

Phenomenology and Mind, n. 29 - 2025, pp. 68-79
DOI: 10.17454/pam-2907
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

I did not know where I was, I did not even understand in the first moment who I was;
I had only, in its original simplicity, the sense of existence
as it may quiver in the depths of an animal (Proust, 1913/2003, p. 7).

1. Introduction

Individuals like you and me are conscious subjects, except when in a coma or dreamless sleep. They, thus, undergo experiences such as seeing, hearing, touching, feeling, and thinking. But, in experiencing, are human persons also invariably aware of their own existence? Does the phenomenology of human subjecthood display a constant, though elementary, sense of ourselves?

William James gave a positive answer to this query. "Whatever I may be thinking of," he claimed in his (1892/1985), "I am always at the same time more or less aware of myself, of my *personal* existence." (p. 43) Terminological dissimilarities notwithstanding, many other prominent historical figures in Western Philosophy appeared to share his viewpoint.[1]

In contemporary analytic philosophy of mind, the opposite attitude has long prevailed.[2] Recently, however, an ever-increasing number of scholars working on the nature of consciousness[3] have come to maintain that every conscious state involves the subject's inner awareness of its occurrence (e.g., Janzen, 2008; Kriegel, 2009b; Levine, 2001, 2006; Zahavi &

1 For instance, Locke famously wrote: "Experience then convinces us, that we have an intuitive knowledge of our own existence, and an internal infallible perception that we are. In every act of sensation, reasoning, or thinking, we are conscious to ourselves of our own being" (Locke, 1690/1997, Ch. 9 sec.3). In a similar vein, and with explicit reference to Locke's view, Leibniz claimed that "we know our own existence by intuition (Leibniz, 1765/1996, bk. IV, Ch. 9, sec. 2). See also, among others, Merleau-Ponty (1945/2010, p. 317).

2 The reason for such a prevalent attitude is debatable. Arguably, however, it has at least partially to do with the prominence that discussions in this area have given to the so-called transparency of experience and the elusiveness of the self. The transparency of experience, which is often tracked back to G. M. Moore (1903), is roughly the view that, when introspecting one's own experience, one can only be aware of features of externally located objects (cf. Harman, 1990; Tye, 1995). The elusiveness of the self, instead, goes back at least to Hume (1739/1975) and corresponds roughly to the view that, in introspection, one cannot be aware of oneself (cf. Howell, 2010; Ryle, 1949).

3 By consciousness here, I mean *phenomenal consciousness*, the property mental states have when, and only when they exhibit a so-called phenomenal character, that is, to put it in Nagel's (1973) famous expression, that "there is something it is like for their subject to be in them." Unless otherwise stated, I will use this meaning of consciousness throughout the paper. Correlatively, when I talk about conscious mental states—or experiences—I mean phenomenally conscious states, namely, states exhibiting a phenomenal character.

Kriegel, 2015).[4] This registers a certain turnaround towards the question at issue, for inner awareness can be understood as a way of shaping such a basic awareness of one's individual existence.[5] Furthermore, inner awareness has been understood in this way in the philosophical literature. For instance, Kriegel (2003, p. 13; 2009b, p. 49) explicitly read the aforementioned Jamesian remark in terms of inner awareness. In this paper, I shall endorse this interpretation, thereby focusing on the elementary sense of our existence *qua* the ever-present inner awareness we have of our conscious states.

Much of the current debate on inner awareness has centered around two main issues. The first issue concerns its existence: is inner awareness always involved in experience? The second concerns its (metaphysical) explanation: in virtue of what does inner awareness ultimately obtain? In what follows, I shall assume that inner awareness exists, focusing on the issue of its explanation. More precisely, my aim is to argue against a family of solutions to such an issue that has gained currency in the relevant literature and that, following Guillot (2017), I will group under the label of the "state self-awareness view." According to this view, roughly, what ultimately grounds inner awareness is that conscious states are aware of themselves.[6] I will contend that, in whatever way one understands it, state self-awareness is unsuited to ground inner awareness. Accordingly, one should look elsewhere for a satisfactory explanation of it.

Here is how I proceed. In section 2, I outline the state self-awareness view in more detail, expanding on its central tenet and briefly considering its main variants. In section 3, I will present the problem that afflicts all versions of the state self-awareness view. Finally, in section 4, I strengthen my case by considering and dismissing a possible way out of the problem.

The reader will not overlook the fact that the concern of the present paper stands primarily within the debate on the metaphysics of consciousness. However, it is worth emphasizing that it is also highly relevant for the philosophical discussions on personal identity, as inner awareness repeatedly enters these discussions. For instance, it is widely held that a certain degree of self-consciousness is essential to personhood (cf. Garrett, 1998).[7] Accordingly, if we regard the awareness of one's own existence as the most basic form of self-consciousness one can display (cf. Di Francesco, Marraffa, & Paternoster, 2016, p. 95), inner awareness turns out to be a key feature of persons, either in the sense that it is a defining feature of them or in the sense that it grounds the higher form of self-consciousness which is essential to them. Furthermore, it has been argued that the basic sense of existence accompanying our ordinary experiences is a necessary precondition for the construction of our personal identity—understood here as the kind of identity that serves us to define *who* we are (cf. Gallagher, 2000, 2013; Zahavi, 2005). Overall, thus, exploring the nature of inner awareness promises to shed important light on the nature of persons and their identities.

2. The State Self-awareness View

The state self-awareness view (SSV) asserts that inner awareness—the subject's constant awareness of their own conscious states—is constituted by these states being aware of themselves. More formally, proponents of SSV endorse the following core thesis (CT):

4 More on the features of such an awareness below.

5 Arguably, by being aware of our occurrent experiences, we do not simply exist *qua* experiencing subjects, but we also have a certain feeling of such an existence.

6 Proponents of the state self-awareness view are, among others, Giustina (2024), Kriegel (2009b), McClelland (2013), Williford (2015), and Zahavi (2005, 2020).

7 See Smith (2017, sec. 4.1) for a review of philosophers who consider self-consciousness essential to personhood.

(CT) For any subject S, conscious state M of S, and time t, S is inner aware of M at t in virtue of M's being aware of itself at t.

Various interpretations of CT exist, leading to different versions of SSV. Specifically, subscribers to CT diverge in (i) their explication of inner awareness (their explanandum) and (ii) their qualifications of the suitable state self-awareness (their explanans). This section delves into these differences to set the stage for the discussion to follow. Since the disagreement on (ii) stems from differing views on (i), I will do this by considering which claims about inner awareness SSV theorists share and which they do not. Hopefully, this will also help clarify the overall theoretical framework of SSV.

However, a caveat is needed before proceeding. Some might find SSV suspicious or outright false, for it suggests mental states are *aware* of themselves and, thus, *subjects of awareness*. To avoid misconstrual, it is worth noticing that most SSV proponents do not claim that mental states are aware like creatures are.[8] "Awareness" in this context tends to be used metaphorically, indicating a special reflexive relationship entertained by conscious mental states. With this clarification in mind, we can examine the main claims about inner awareness shared by all subscribers to CT.

The first claim I consider fixes the place of inner awareness within consciousness. As noticed, advocates of inner awareness take it to be present in all conscious experiences, maintaining that one is necessarily aware of it whenever one is in a conscious state. Proponents of SSV craft such a ubiquity claim rather strongly, maintaining that inner awareness is not only *necessary* for consciousness but also *constitutive* of it. Thus, they endorse the claim that [1] inner awareness is what makes a mental state conscious.

What underlies [1] is the view that conscious mental states essentially feature so-called *for-me-ness*, the conscious states' property of being "for their subject" (of being phenomenally given to their subject).[9] This property, in fact, is typically spelled out in terms of inner awareness (cf. Janzen, 2008, p. 156; Kriegel, 2009b, pp. 47-48). Moreover, it is taken to constitute consciousness *as such*—what makes a mental state conscious. On this view, thus, inner awareness is the *source* of consciousness, what makes a conscious state for me and, thereby, conscious in the first place.

[1] is a claim SSV shares with other accounts of inner awareness, most notably higher-order monitoring accounts (see, among others, Coleman, 2015; Gennaro, 2012; Lycan, 2001; Rosenthal, 1997), from which SSV derives. These accounts also ground inner awareness in an awareness relation borne by mental states. However, while SSV views this as a relation conscious states bear to themselves, higher-order monitoring accounts see it as a relation between two distinct mental states, where one state is directed at another.

SSV theorists depart from higher-order monitoring accounts based on three additional claims they share about inner awareness. First, [2] inner awareness is phenomenal, meaning we are experientially presented with our conscious states; we are conscious of them. Second, [3] inner awareness is non-reflective. While reflective awareness is introspective, objectifying (viz., presenting the object of awareness as standing in opposition to the subject), conceptual, and possibly inferential, inner awareness is not. Third, [4] inner awareness puts the subject in an intimate relationship with their conscious states, meaning our awareness of our conscious

8 I say "most proponents" because some versions of the view—specifically some variants of the non-representational account—arguably imply that mental states are literally aware of themselves.

9 See Zahavi & Kriegel (2015) for a detailed discussion of for-me-ness as an essential component of the phenomenal character of conscious mental states.

states cannot be illusory or hallucinatory, unlike our perception of external objects. SSV theorists argue that higher-order monitoring accounts struggle to accommodate these claims, while SSV easily integrates them, making it a more plausible account (Janzen, 2008; Kriegel, 2009b; Zahavi, 2004).

Other claims about inner awareness are more controversial than those just mentioned among SSV theorists. Two claims, in particular, cause the main divisions within the SSV family tree. The first is that [5] inner awareness involves the subject *representing* its conscious state. Subscribers to [5] maintain that the "of" of inner awareness is the "of" of intentionality. Accordingly, they unpack the state self-awareness suitable to ground inner awareness in representational terms, advocating for *self-representationalism*:

(SR) for any conscious state M of a subject S, S is inner aware of M (and hence M is for me) in virtue of M's suitably representing itself.

Self-representationalism comes in different varieties, depending on how the state self-representation is qualified. One influential variety qualifies it as specific, essential, and non-derivative (Kriegel, 2009b). However, these details are not crucial for what follows. What matters is that, according to the self-representationalist branch of SSV, a conscious state always suitably represents itself, making it a state of which the subject is inner aware.

Some SSV theorists deny [5] instead on the basis that the intimacy of inner awareness speaks against it. Accordingly, they construe the relevant state self-awareness in non-representational terms, typically in terms of the acquaintance relation, defending thus *self-non-representationalism*:

(SNR) for any conscious state M of a subject S, S is inner aware of M (and hence M is for me) in virtue of M's bearing a suitable non-representational relation to itself.

The exact nature of this putative non-representational relation is debated. For instance, Zahavi (2005, 2014, 2020) and Williford (2015, 2019) explicate it in terms of a primitive acquaintance relation conscious states bear to themselves. Giustina (2024), instead, appeals to self-quotation, understood as the mental state's relation of quoting or displaying itself.

The second main controversial claim among SSV theorists is that [6] inner awareness involves the awareness of the subject themself in addition to their conscious states. [6] concerns the object of inner awareness. Its subscribers diverge in how they take the subject to figure in inner awareness. In a modest view, the subject appears merely alongside the experience. In a more demanding view, the subject appears as the owner of the experience.[10] Regardless of these differences, [6] marks out the distinction within the SSV family between what—following Sebastián (2012)—we might call subject-involving (SI) and mental state-involving (MSI) accounts of inner awareness:

(SI) for any conscious state M of a subject S, S is inner aware of M (and hence M is for me) in virtue of M's being somehow aware of S and itself.

(MSI) for any conscious state M of a subject S, S is inner aware of M (and hence M is for me) in virtue of M's being just aware of itself.

10 For more on these views, see Guillot (2017)

Such a distinction is orthogonal to the one between self-representationalism and non-self-representationalism. These views, thus, admit both subject-involving and mental state-involving versions. For instance, Kriegel (2009b) defends a mental state-involving version of self-representationalism, while Sebastián (2012) proposes a subject-involving variant. With a clearer understanding of the SSV family tree, we can now turn our attention to what, I believe, makes it a fruitless tree.

3. The Problem of *State* Awareness

SSV has been questioned on several grounds. Some critiques have focused on specific versions of the view, while others have taken a broader approach, questioning the core of SSV. The latter chiefly revolves around either SSV's account of the explanandum (cf. Dainton, 2016; Howell & Thompson, 2017) or the explanatory adequacy of the state self-awareness, specifically whether a reflexive awareness of mental states can account for them being objects of inner awareness (cf. Gennaro, 2012, ch. 5).

One aspect, however, has not been the subject of much criticism: that inner awareness ultimately depends on an awareness relationship entertained by mental states. Put another way, what is not seen as particularly problematic within the SSV theoretical framework is that my inner awareness of my occurrent experience is ultimately grounded in a state awareness, whether or not it takes the form of self-awareness.[11] I find this attitude misleading. In my view, the deepest issue of SSV lies precisely in this "state-first" approach to the explanation of inner awareness—in its attempt to account for inner awareness through an awareness relationship held by mental states.

My concern develops upon a criticism Joseph Levine (2006) raised against the self-representationalist branch of SSV, specifically against its mental state-involving variant. This view, I recall, accounts for inner awareness in terms of a mental state's suitable representation of itself. Since inner awareness constitutes the for-me-ness of conscious states—what makes a mental state conscious—this is tantamount to claiming that mental states are for their subject—and hence conscious—in virtue of being suitably self-representing states.[12]

According to Levine, it is unclear how self-representing states could recover the for-me-ness of our conscious states. As long as self-representing states represent themselves in the same way they represent other things, he claims, one can reasonably ask what makes such states for me: states of which I am aware. As he puts it,

> Somehow, what we have in conscious states are representations that are intrinsically of subjective significance, "animated" as it were, and I maintain that we really don't understand how that is possible. It doesn't seem to be a matter of more of the same—more representation of the same kind—but rather representation of a different kind altogether (Levine, 2006, p. 195).

Such a criticism is sometimes considered one of the most compelling objections to self-representationalism. Kriegel (2011)—a prominent defender of such an account—considers it "the deepest objection to self-representationalism" (p. 69). On this ground, some scholars have appealed to it to support self-non-representationalism.[13] However, on scrutiny, Levine's

11 To be sure, I am not saying that nobody has ever questioned this aspect. In fact, I will point out in a moment a criticism Levine raised against self-representationalism that goes in this direction, which, for this reason, I will elaborate on. As we shall see, however, this criticism is not typically intended to undermine the whole SSV family.
12 The most well-developed model of this view is due to Kriegel (2005; 2009a; 2009b; 2011).
13 See, for instance, Giustina (2022).

criticism is much broader in scope, running over the whole SSV family tree.[14] Or, at least, adapting it to this purpose is relatively easy. To appreciate this, it is crucial to pinpoint the heart of the objection.

The bulk of Levine's objection is that suitable self-representations cannot account for mental states' being for us because they merely add more representations of the kind involved in representing other things. But why are the latter representations unsuited to recover inner awareness/for-me-ness? I contend that the issue lies not in what mental states represent or how they represent but in the fact that *they* represent—that mental states are the ultimate subject of the representation relation. In other words, the difficulty in understanding why a state representing itself suffices to make it "not merely something happening within her, but 'for her'" (Levine, 2006, p. 195) stems from its being a *state representation*, just like other familiar kinds of representations.

Indeed, understanding how a particular state self-representation can explain the subject's inner awareness of their mental state is puzzling. The explanandum refers to a relation between the subject and their mental state, while the explanans refers to a relation between the mental state and itself. Given that, one might reasonably ask why and how the latter can recover the former. Even if a specific qualification of the right sort of self-representation were extensionally adequate, we would still face an explanatory gap between for-me-ness and this relevant self-representation.

Significantly, in considering Levine's complaint, Kriegel (2011) provides a similar analysis of its core issue. "What Levine's line of objection seems to press," he claims, "is the need for a sui generis notion of representation-for-me, *a kind of primitive intentional relation borne by subjects*, rather than by subjects' internal states" (*Ibid.*, p. 19, my emphasis).

However, if this is the case, it is evident that such a problem outstrips the boundaries of self-representationalism, for it does not really concern the representational treatment such an account gives to the state self-awareness suitable to ground inner awareness. Rather, it concerns the more general claim that we can account for inner awareness by appealing to a mental state's relationship. As such, the question Levine raises to self-representationalism can be raised—with equal force—to other variants of SSV. Consider, for instance, self-non-representationalism in its mental state-involving variant, which holds that inner awareness is ultimately grounded in a non-representational awareness that a subject's state has of itself. Just as with self-representing states, we can legitimately ask why a mental state of us being non-representationally aware of itself makes it a state of which we are inner aware: not just 'in us' but 'given to us.'

Qualifying the right kind of state non-representational awareness does not help much in addressing such a question. Whether we construe it as the mental state's relation of quoting (or displaying) itself (Giustina, 2024) or as the mental state's relationship of being acquainted with itself (Zahavi, 2014; 2020; Williford, 2015; 2019), we still have no grip on why this should have anything to do with my inner awareness of the state. There is daylight between it and the relationship it is supposed to explain. Inner awareness intimately relates me to my conscious state, while the putative non-representational relationship invoked to explain it, at best, intimately relates my mental state to itself.

Including myself among the objects of the relevant non-representational awareness also does not help. That is, even if we endorse a subject-involving variant of self-non-

14 The instrumental use that some proponents of the non-representational variant of SSV make of Levine's critique demonstrates how unproblematic it is deemed to ground inner awareness in state awareness.

representationalism, the relevant awareness relation remains at the wrong level, so to speak; it is not a relation borne by me but by an internal state of mine.[15]

Overall, a deep explanatory gap lies at the roots of the SSV family tree. An account of inner awareness seeks to explain what makes us aware of our conscious states, what makes the latter not just 'in us' but 'for us' (conscious in the first place). We are seeking what grounds a subject-level phenomenon. However, it is hard to see how state self-awareness could fulfill this role, as it is a mental state-level phenomenon. All this kind of awareness might reasonably recover is that a mental state is given to itself, namely, that it instantiates the property of for-*itself*-ness. This is the wrong kind of property instantiation: the property of conscious states we seek is for-*me*-ness. Accordingly, no matter how one specifies it, for any state self-awareness taken to ground inner awareness, the following gap-revealing question stands out: why and how do we get inner awareness out of it? Why and how does for-me-ness obtain in virtue of such a for-itself-ness? I call this the problem of state awareness for SSV.

4. A (unsuccessful) way out

I believe the problem of state awareness is the deepest issue of SSV, for it questions the very intelligibility of the view. Without an explanation of why and how the grounding relation between state self-awareness (the grounding fact) and inner awareness (the grounded fact) holds, we cannot fully grasp what it means for the latter to obtain in virtue of the former.

It might be objected, however, that I failed to consider a thesis endorsed by many SSV theorists that effectively backs SSV with an explanation of the relevant grounding relationship, thereby dissolving the problem at hand. This is to contend, in other words, that the explanatory gap, on which my issue rests, exists, but it is harmless, as it is just *apparent*. The thesis I have in mind is what we might call the *priority of state consciousness* (PSC) thesis.[16] Let us examine it more closely.

Since Rosenthal (1986) it is customary to distinguish between consciousness as a property of subjects, expressed by sentences like "I am conscious" or "I am conscious of the tree in front of me," and consciousness as a property of mental states, expressed by sentences like "Marco's desire is conscious." The former is called creature consciousness, and the latter is called state consciousness. These properties are further specified by drawing the parallel distinction between *transitive* and *intransitive consciousness*. The former is a relational property expressed by the transitive locution 'x is conscious of y.' The latter is an intrinsic property expressed by the intransitive locution 'x is conscious.' Accordingly, four types of consciousness are typically distinguished conceptually: *intransitive creature consciousness, transitive creature consciousness, intransitive state consciousness*, and *transitive state consciousness*. The latter may sound odd to some readers, but like state awareness, the use of "conscious of" here is idiosyncratic. Transitive state consciousness is a *term of art*, referring to the property mental states have when, and only when, their subjects are transitively conscious of something in virtue of being in them (cf. Kriegel, 2009, p. 28).

A natural question is what relationship, if any, exists among these four properties. PSC addresses such a question. It holds that intransitive state consciousness, the mental state's property of *being conscious*, is explanatorily prior to the other three properties, in the sense that the latter all ultimately depend on the former.

PSC is widely endorsed in the literature on consciousness, though few philosophers

15 The same reasoning applies to the subject-involving variant of self-representationalism.

16 The thesis identifying subjects of experience with their experiences might be equally relevant. Space constraints, however, forced me to make a choice on which thesis to consider. I decided to discuss PSC because the other is somehow marginal in the literature (possibly endorsed just by Zahavi and Williford) and highly revisionary (Barbieri 2025; Bayne, 2008).

express it explicitly, and even fewer argue for it.[17] To my knowledge, Kriegel (2003, 2009b) is the only one who makes a case for it among SSV theorists. He argues that both kinds of creature consciousness depend on their state consciousness counterparts. Thus, a subject is intransitively conscious in virtue of being in an intransitively conscious state and transitively conscious in virtue of being in a transitively conscious state. Moreover, he argues that transitive state consciousness depends on intransitive state consciousness. According to him, in fact, a mental state cannot be transitively conscious—making its subject transitively conscious of something—without being intransitively conscious in the first place (*Ibid.,* pp. 30-31).

PSC seems to solve the problem of *state* awareness straightforwardly. The subject's inner awareness of their conscious state is a conscious phenomenon and, to that extent, a form of transitive creature consciousness. Since the latter depends on transitive state consciousness, it is natural to view state self-awareness as a kind of transitive state consciousness: the property a mental state has of making its subject conscious of itself. Thus, PSC backs SSV with an explanation of why and how the putative grounding relation holds: inner awareness is obtained in virtue of state self-awareness *because* (transitive) state consciousness is prior to (transitive) creature consciousness.

Unfortunately, this straightforward explanation for backing SSV does not work, for PSC is in tension, if not plainly inconsistent, with [1], the claim that inner awareness is constitutive of consciousness.[18] According to SSV, recall, the subject's inner awareness of their occurrent mental state *makes* the latter conscious *simpliciter*, that is, intransitively conscious. This means that (a form of) transitive creature consciousness is metaphysically prior to intransitive state consciousness: a mental state is intransitively conscious *in virtue of* its subject being conscious of it. However, since PSC claims that transitive state consciousness ultimately depends on intransitive state consciousness, it follows that the suitable awareness my occurrent conscious state has of itself would be both what *ultimately grounds* my inner awareness and what *is ultimately grounded* in my inner awareness. Thus, not only does PSC lack the resources to make the grounding relationship advocated by SSV intelligible, but it also adds a problematic explanatory circularity to it. Overall, appealing to PSC to way out from the problem of *state* awareness is a dead end.

5. Conclusion

In contemporary philosophy of mind, the sense of personal existence that many classical philosophers have believed to permeate our experience is often cashed out in terms of the ubiquity of one's inner awareness of one's own experience. In this paper, I addressed what explains this inner awareness, arguing against the view that it is ultimately grounded in a fundamental awareness that my occurrent experience has of itself. This is the state self-awareness view (SSV).

I began by detailing the core of SSV and its main variants. I then showed that all versions of SSV suffer from the problem of *state* awareness, which is the problem of understanding why and how a mental state-level relation like state self-awareness could recover a subject-level relation like inner awareness. Finally, I defended the severity of this problem for SSV by blocking a potential reply.

Naturally, my take on the explanatory issue of inner awareness has been more negative than positive: I argued for what does not ground inner awareness rather than what grounds

17 Philosophers who formulate PSC explicitly include, among others, Block (1995), Dretske (1993), Gennaro (2012), Kriegel (2009b), and McClelland (2012).

18 I discussed more extensively the inconsistency at hand in (Barbieri, 2024).

it. However, eliminating some accounts from the range of possibilities is already a significant achievement, especially when these accounts foreground the discussion. Furthermore, I believe that the problem of *state* awareness points strongly to the fact that a promising account of inner awareness cannot dispense with treating it fundamentally as it appears: a relation entertained by the subject rather than their mental states. However, defending such a claim in detail and elaborating on the nature of such a subject's relationship is an additional task, a task for another work.

REFERENCES

Barbieri, A. (2024). Taking Phenomenology at Face Value: The Priority of State Consciousness in Light of the For-me-ness of Experience. *Argumenta*, 9(2), 517-538. doi:10.14275/2465-2334/20230.bar;

Barbieri, A. (2025). What Is Pre-Reflective Self-Consciousness an Awareness of? An Argument for the Egological View. *The Southern Journal of Philosophy*, 63(4), 533-552. https://doi.org/10.1111/sjp.12614;

Bayne, T. (2008). The Phenomenology of Agency. *Philosophy Compass*, *3*(1), 182-202. https://doi.org/10.1111/j.1747-9991.2007.00122.x;

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227-287. https://doi.org/10.1017/S0140525X00038188;

Coleman, S. (2015). Quotational higher-order thought theory. *Philosophical Studies 2015 172:10*, *172*(10), 2705-2733. https://doi.org/10.1007/S11098-015-0441-1;

Dainton, B. (2016). I—The Sense of Self. *Aristotelian Society Supplementary Volume*, *90*(1), 113-143. https://doi.org/10.1093/arisup/akw007;

Di Francesco, M., Marraffa, M., & Paternoster, A. (2016). *The Self and its Defenses: From Psychodynamics to Cognitive Science*. London: Palgrave Macmillan;

Dretske, F. (1993). Conscious Experience. *Mind, New Series*, *102*(406), 263-283;

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14-21. https://doi.org/10.1016/S1364-6613(99)01417-5;

Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience*, *7*(August), 1-7. https://doi.org/10.3389/fnhum.2013.00443;

Garrett, B. (1998). *Personal Identity and Self-Consciousness*. New York, NY: Routledge;

Gennaro, R. J. (2012). *The Consciousness Paradox: Consciousness, concepts, and High-Order Thoughts*. Cambridge, MA: MIT Press;

Giustina, A. (2022). An Acquaintance alternative to Self-Representationalism. *Philosophical Studies*, 1-32. https://doi.org/10.1007/S11098-022-01868-5;

Giustina, A. (2024). Inner Acquaintance Theories of Consciousness. In U. Kriegel (Ed.), *Oxford Studies in Philosophy of Mind: Volume 4* (p. 0). Oxford: Oxford University Press. https://doi.org/10.1093/9780198924159.003.0010;

Guillot, M. (2017). I Me Mine: On a Confusion Concerning the Subjective Character of Experience. *Review of Philosophy and Psychology*, *8*(1), 23-53. https://doi.org/10.1007/s13164-016-0313-4;

Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives*, *4*, 31-52. https://doi.org/10.2307/2214186;

Howell, R. J. (2010). Subjectivity and the Elusiveness of the Self. *Canadian Journal of Philosophy*, *40*(3), 459-483;

Howell, R. J., & Thompson, B. (2017). Phenomenally Mine: In Search of the Subjective Character of Consciousness. *Review of Philosophy and Psychology*, *8*(1), 103-127. https://doi.org/10.1007/s13164-016-0309-0;

Hume, D. (1975). *A Treatise of Human Nature* (2nd ed.). Ed. L. A. Shelby-Bigge. Oxford: Clarendon Press (Original work published 1739);
James, W. (1985). *Psychology: The Briefer Course*. Ed. G. Allport. Paris: University of Notre Dame Press (Original work published 1892);
Janzen, G. (2008). *The Reflexive Nature of Consciousness*. Amsterdam: John Benjamins Publishing Company, John;
Kriegel, U. (2003). Consciousness as sensory quality and as implicit self-awareness. *Phenomenology and the Cognitive Sciences*, *2*(1), 1-26. https://doi.org/10.1023/A:1022912206810;
Kriegel, U. (2005). Naturalizing Subjective Character. *Philosophy and Phenomenological Research*, *LXXI*(1);
Kriegel, U. (2009a). Self-representationalism and phenomenology. *Philosophical Studies*, *143*, 357-381. https://doi.org/10.1007/s11098-008-9204-6;
Kriegel, U. (2009b). *Subjective Consciousness*. New York, NY: Oxford University Press;
Kriegel, U. (2011). Self-Representationalism and the Explanatory Gap. In J. Liu & J. Perry (Eds.), *Consciousness and the Self: New Essay* (pp. 51-75). Cambridge: Cambridge University Press;
Leibniz, G. W. (1996). *New Essays on Human Understanding* (2nd ed.). Trans. P. Remnant & J. Bennett. Cambridge: Cambridge University Press (Original work published 1765);
Levine, J. (2001). *Purple Haze. The Puzzle of Consciousness*. Oxford: Oxford University Press;
Levine, J. (2006). Conscious Awareness and (Self-)Representation. In U. Kriegel & K. Williford (Eds.), *Self-Representational Approaches to Consciousness* (pp. 173-198). Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/6155.003.0010;
Locke, J. (1997). *An Essay Concerning Human Understanding*. Ed. R. S. Woolhouse. London: Penguin Classic (Original work published 1690);
Lycan, W. G. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis*, *61*(1), 3-4. https://doi.org/10.1093/analys/61.1.3;
McClelland, T. (2013). The Neo-Russellian Ignorance Hypothesis: A Hybrid Account of Phenomenal Consciousness. *Journal of Consciousness Studies*, 20(2-4), 121-151;
Merleau-Ponty, M. (2010). *Phenomenology of Perception*. Trans. D. A. Landes. London: Routledge (Original work published 1945);
Moore, G. E. (1903). The Refutation of Idealism. *Mind*, *12*(48), 433-453;
Proust, M. (2003). *In Search of Lost Time Vol. 1: The Way by Swann*. Ed. C. Prendergast. Trans. L. Davis. London: Penguin Classic (Original work published 1913);
Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, *49*(3), 329-359. https://doi.org/10.1007/BF00355521;
Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. J. Flanagan, & G. Guzeldere (Eds.), *The Nature of Consciousness* (pp. 729-753). Cambridge, MA: MIT Press;
Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson;
Smith, J. (2017). Self-Consciousness. In *Stanford Encyclopedia of Philosophy* (p. 89);
Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press;
Williford, K. (2015). Representationalism, Subjective Character, and Self-Acquaintance. In T. Metzinger & M. Windt (Eds.), *Open MIND*. Frankfurt am Main: MIND Group;
Zahavi, D. (2004). Back to Brentano. *Journal of Consciousness Studies*, *11*(10-11), 66-87;
Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: Bradford Book/MIT Press;
Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. London: Oxford University Press;
Zahavi, D. (2020). Consciousness and (minimal) selfhood: Getting clearer on for-me-ness and

mineness. In U. Kriegel (Ed.), *The Oxford Handbook of philosophy of consciousness* (pp. 1-19). Oxford: Oxford University Press;

Zahavi, D., & Kriegel, U. (2015). For-me-ness: What it is and what it is not. In D. Dahlstrom, A. Elpidorou, & W. Hopp (Eds.), *Philosophy of Mind and Phenomenology* (pp. 36-53). New York, NY: Routledge.

ALFREDO TOMASETTA
IUSS Scuola Universitaria Superiore Pavia
alfredo.tomasetta@iusspavia.it

MULTIMODALITY AND NO-SELF

abstract

The Indian Buddhist philosopher Vasubandhu argued forcefully for the no-self view. Philosophers belonging to the Hindu Nyāya tradition disagreed: the self is required, among other things, to explain the existence of multimodal perceptions. Monima Chadha has recently responded to this multimodality argument on behalf of Vasubandhu. In this paper I argue that her response is problematic in two ways: it is not consistent with Vasubandhu's general ontology, and it proposes an unstable position.

keywords

no-self, multimodality, Vasubandhu, Nyāya, Chadha

Phenomenology and Mind, n. 29 - 2025, pp. 80-89
DOI: 10.17454/pam-2908
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Philosophers in the Hindu Nyāya tradition who oppose the Buddhist no-self thesis have offered various arguments for the existence of the self. Notable among these is the idea that the existence of multimodal experience requires a self capable of synthesising different sensory data. Against Nyāya's multimodality argument, Monima Chadha, in her recent book *Selfless Minds*, offers a defence of the no-self thesis as elaborated by the Buddhist philosopher Vasubandhu (Chadha, 2023). She claims that there are indeed multimodal perceptual experiences, but that they do not require a self. In this article I argue that this defence is problematic. One reason is that it conflicts with Vasubandhu's ontology. But even leaving this problem aside, I suggest that Chadha's position either paves the way for the admission of the self as a basic entity, or ends up renouncing multimodal experiences.

Here is the outline of the paper. Section 2 gives an overview of Vasubandhu's no-self thesis. Section 3 presents the multimodality argument for the existence of the self put forward by Nyāya philosophers and Chadha's response to this argument on behalf of Vasubandhu. In section 4 I argue that this response is problematic for the reasons just suggested. Finally, section 5 briefly summarises the discussion.

2. Vasubandhu's no-self view

Let us begin, then, with a (very sketchy) presentation of the doctrine of no-self as elaborated by Vasubandhu.

Vasubandhu is a Buddhist philosopher who lived between the 4th and 5th centuries CE, and wrote several works, including the monumental *Treasury of Metaphysics*.[1] In particular, in the ninth book of this work, Vasubandhu presents various arguments against the existence of the 'I', or the self.

The 'I', says Vasubandhu, does not exist: neither as a soul, nor as a person in the ordinary sense, i.e. as a unified psycho-physical complex that persists over time.

Vasubandhu denies the existence of the soul because, he claims, it cannot be known. There are three reliable means of knowledge, he says (in accordance with the philosophical tradition to which he himself belongs): perception – external and internal –, inference, and authoritative texts. None of them, Vasubandhu argues, supports the existence of the soul.[2]

The denial of ordinary persons – as I said: unified psycho-physical complexes that persist

1 In Sanskrit, the *Abhidharmakośabhāṣyam*.

2 On this topic see Kellner & Taber (2014).

over time – depends on some radical theses that are argued in a very subtle way. I will not go into these arguments here, and will content myself with stating the theses.[3]

Thesis A)
There are no substances, i.e. there are no property-bearers.

Thesis B)
Mereological nihilism is true: there are no entities composed of parts, only simple entities.

Thesis C)
Momentariness: nothing lasts longer than a moment.

Thesis D)
Reality consists only of *dharmas.*

"*Dharma*" is a Sanskrit keyword that can be translated in this context as "property" and also as "fundamental element".
In particular, *dharmas* are occurrences of properties (tropes), simple entities (i.e. without parts), either mental or physical, of momentary duration, and causally dependent on other *dharmas*.
This particular wetness or this particular solidity are examples of physical *dharmas*; this particular pain or this particular perception of colour are examples of mental *dharmas*.

Vasubandhu's metaphysics, as seen in the four theses just presented, crumbles the world of common sense, both synchronically and diachronically. For common sense, there are complex entities that persist over time (pots, trees, people...). For Vasubandhu there are only simple entities, the *dharmas*. They are occurrences of physical or mental properties of instantaneous duration, causally related to each other. There are no common-sense entities, such as pots or persons; when one thinks one is dealing with such things, what actually exists are misrepresentations ('pot', 'person'): these misrepresentations are produced by conceptualising, as a synchro-diachronic unit, successions of simple and instantaneous entities which are precisely the *dharmas*. As Vasubandhu says: "Instantaneous conditioned factors [i.e. *dharmas*] whose succession is unbroken are considered by the ignorant as 'Devadatta' by taking a group as one being" (AKBh, p. 476).[4,5]

One final remark. The metaphysics that I have sketched out very briefly is a means to the soteriological end of Buddhism. The goal of Buddhism is liberation from suffering, the first cause of which is ignorance; and the most serious ignorance is to assume the existence of a unified and persistent self/ego through time. This is a dangerous fiction created by mistaking certain multiplicities of simple and instantaneous *dharmas* for a synchro-diachronic unity. Metaphysical enquiry is necessary to free oneself from this error, which is the primary source of suffering, says Vasubandhu. Metaphysics will set you free, one might say.

3 But see Gold (2022), Siderits (2021, pp. 150-203), Tomasetta (2025, pp. 183-249), and Watson (2017).
4 "Devadatta" is a common Sanskrit name for a person.
5 The translation is mine. See also, Kapstein (2001, pp. 368-369).

3. Nyāya multi-modality argument for the self, and Chadha's reply

While the no-self thesis is something of a hallmark of Buddhism, the assertion of the existence of the self is characteristic of most non-Buddhist Indian philosophies. Here I focus on the Nyāya school ("Nyāya" means "method" or "logic"), which is about two thousand years old and still alive today.[6]

One of the traditional arguments for the existence of the self offered by the Nyāya philosophers is based on multimodal perception: the self is needed to explain the possibility of multimodal experience. Let us see.

When we experience a character speaking in a film, we see her mouth move and hear her words: the visual and auditory data appear to be parts of a single complex visual-auditory experience, a so-called multimodal experience, an experience that consists of perceptual data provided by two different sensory modalities, vision and hearing.

The canonical example of multimodal perception offered by the Nyāya philosophers – which, of course, does not refer to cinema – is that of a visuo-tactile experience: I now touch what I see – I feel with my hands the roughness of the fruit I am observing, let us say. In this case, the visual and tactile data seem to be part of a single multimodal, visuo-tactile perceptual experience.

But one might ask: how is such a complex experience possible? According to the Nyāya philosophers, the unity of this multimodal experience can only be explained by appealing to a conscious entity, a self, that is able to *synthesise* the different sensory data provided by touch and sight into a single experience.

In other words, the manifest fact of the existence of multimodal experience (assuming it is indeed a fact) needs to be explained. For the Nyāya philosophers, the only way to explain this fact is to admit an entity capable of synthesising the perceptual data provided by different sensory modalities into a single experience. This entity, this principle capable of synthesising the data provided by different senses, is the self or the 'I'. The explanation of the unity of certain complex experiences therefore requires the existence of the self/ego/I. As Vātsyāyana writes:[7]

> Some particular object is grasped by sight; the same object is also grasped by touch. "That very thing which I saw with my eyes, I am now feeling through my sense of touch", and "That very thing which I felt through my sense of touch, I am now seeing with my eyes [...]. These two perceptions of a single object are synthesised by a single agent [...] Thus the grasper of one and the same object [...] synthesises two perceptions about one and the same thing; [...] That one, in a special category, is the self (NSBh, p. 135/Dasti & Phillips, 2017, pp. 80-81, slightly modified).

One must therefore postulate the self as the entity that, by synthesising different perceptual data, explains the existence of multimodal perceptions.

Now, the multimodality argument for the existence of the self we have just considered was made by Nyāya philosophers who lived *after* Vasubandhu, and this explains why there is no reply in Vasubandhu's texts. In her recent book *Selfless Minds*, Monima Chadha attempts, among other things, to respond on behalf of Vasubandhu to the argument for the existence of the self based on multimodal perceptions (Chadha, 2023, pp. 98-118).[8]

6 For accessible introductions see Dasti (2023) and Dasti & Phillips (2017).

7 Vātsyāyana (5th century CE) is one of the greatest philosophers of the Nyāya tradition.

8 Chadha's focus is really on the phenomenon of what she calls "synchronic unity". This certainly includes multimodal perceptions, but also such supposedly complex perceptions as seeing the table and the sofa at the same time (Chadha, 2023, p. 108), experiences that are obviously not multimodal perceptions. Since I believe that apparently

According to Chadha, we should indeed admit that multimodal perceptions – henceforth "MMPs" – exist. The point, she argues, is that there is no need for a synthesising self in order to have MMPs: visual and tactile (or other perceptual) data combine on their own to produce complex experiences.

This admission, Chadha insists, does no harm to Buddhist soteriology (Chadha, 2023, pp. 115-117). What is liberating for Buddhists is to know that there is no self; but the admission of multimodal perceptions does not threaten the no-self doctrine. There is no soul, and there are no ordinary persons: there are only successive momentary physical and mental *dharmas*; it is just that some of the mental *dharmas* form wholes made up of parts (these wholes are the MMPs); and that is all.

4. Two problems for Chadha's reply

We have just considered Chadha's response to Nyāya's multimodality challenge. I think there are at least two problems with this response. The first problem is that what Chadha says does not agree with what Vasubandhu himself says about what exists – and since Chadha is responding on behalf of Vasubandhu this is significant. The second problem is that hers is an unstable position, in a sense that will be clarified. Let us start with the first problem.

4.1 Disagreement with Vasubandhu

According to Chadha, MMPs are to be allowed. And MMPs are, of course, wholes in the mereological sense. But we have seen that, according to Vasubandhu, mereological nihilism is true: there are no wholes composed of parts, only simple entities. So the existence of MMPs is obviously incompatible with Vasubandhu's position. End of story.

Is it that simple? Here is a possible rejoinder. In Vasubandhu's text there is no explicit affirmation of the *full generality* of the thesis that wholes do not exist. While there may be reasons to attribute a generalised mereological nihilism to Vasubandhu, this reading is not entirely uncontroversial. Exegetically, that is, it is not obvious that, for Vasubandhu, there are only simple entities. It could be argued that, for him, there are at least cases of composite entities, and these cases coincide with the MMPs.

Against this answer, I offer two considerations designed to show that the textual data most likely rule out a non-generalised mereological nihilism.

First note that Vasubandhu distinguishes between basic entities and entities 'by convention', and that this distinction is exhaustive. The basic entities are the *dharmas,* which are without parts (see AKBh, p. 334/Siderits, 2021, pp. 163-165). So MMPs, which have parts, are not basic entities. Since the distinction between basic entities and conventional entities is exhaustive, MMPs must be conventional entities.

Now, conventional entities include substances, past and future entities, absences and persons. And Vasubandhu says that these things do not exist – indeed, Vasubandhu's denial of the existence of persons is one of the main theses of Chadha's reading of Vasubandhu himself.[9]

Paradigmatic 'conventional entities' – such as substances, past and future things, absences and persons – are therefore, somewhat paradoxically, not entities. But it is not difficult to give an explanation for this: a fake painting by Rembrandt is not a painting by Rembrandt that is a fake; it is not a painting by Rembrandt. Similarly, a conventional entity is not an entity that is conventional: it is not an entity.

complex *mono-modal* perceptions raise quite different issues that deserve separate treatment, I will limit my discussion to multimodality – a move that is also justified by the fact that this is the only focus of the Nyāya argument under consideration here.

9 For the cases of substances, absences, and past and future entities, see, for example, Gold (2015, pp. 22-58), Guerrero (2023), and Watson (2017).

So, conventional paradigmatic entities are not entities: they are reduced to names/concepts that are conventionally used in the presence of certain *dharmas*. For example, the word/concept 'person' – which purports to refer to a unified psycho-physical complex that persists over time – is used in the presence of certain organised 'person-wise' tropes. But if conventional paradigmatic entities are not entities, it is hard to see why MMPs, which are conventional entities, should be real entities instead. Vasubandhu does not explicitly mention any exception in this regard, which is odd, if he really believes that MMPs are real (conventional) entities.

What's more, and this is my second brief consideration, Vasubandhu explicitly says that wholes, such as a carpet or a cloth, have no reality beyond the linguistic-conceptual one (see AKBh, 189/Siderits, 2021, p. 156). And he makes no explicit exception for perceptions. Which, again, is very strange, if he really believed that MMPs are real wholes.

To sum up. MMPs are incompatible with Vasubandhu's mereological nihilism, at least as it is usually understood, as the negation, that is, of the existence of all wholes. But the possibility that the usual interpretation of nihilism is incorrect because of the existence of MMPs does not seem to be exegetically tenable. Then, if Vasubandhu's mereological nihilism is indeed entirely general, Chadha's admission of MMPs is incompatible with what Vasubandhu says about what exists.

4.2 Chadha's position is unstable

Mereological wholes such as MMPs should be allowed to exist, says Chadha. But once set in motion, a snowball can cause an avalanche: if there are mental wholes of a certain kind, why not admit other wholes as well, even to the point of admitting those psycho-physical synchro-diachronic units that are ordinary people? Except that Chadha, along with Vasubandhu, certainly does not want to admit that the self exists, even in this ordinary version.

So, to stop this slippery slope – the slippery slope that leads from the admission of MMPs to the existence of wholes like ordinary people – Chadha needs an 'asymmetrical mereology', so to speak:

> Yes, there are MMPs – a certain kind of whole.
>
> No, there are no other wholes – in particular, there are no (ordinary) persons.

Chadha seems to me to suggest two strategies for justifying the yes/no answer I have just formulated, and thus for stopping the slippery slope.[10] I will now consider these two strategies and argue that they are both problematic. This will lead to the conclusion that Chadha's position is unstable.

Strategy 1. Phenomenology

Chadha talks about "the manifest fact that conscious experiences often seem to be composed of multiple features, multiple modalities". This is, she says, "the manifest fact of synchronic unity" (Chadha, 2023, p. 101). So, there is phenomenological evidence for MMPs: that there are MMPs seems to be a manifest phenomenological fact – MMPs appear to exist, and to exist as unified experiences composed of parts.

10 Strictly speaking, they are two strategies *for admitting MMPs*; but it is clear that Chadha thinks they do not work for admitting the existence of other wholes (and in particular persons in the ordinary sense). I add that Chadha does not explicitly isolate these two strategies, but I take them to be her basic line of argument.

My reply to this phenomenological strategy begins with the following consideration: there also appears to be a unity of conscious experience as a whole at any given time; the experiences, the conscious mental states, appear as parts of a single phenomenological field. In this regard Bayne writes:

> Consider the structure of your overall conscious state. I suspect that you will be inclined to the view that all your current experiences are phenomenally unified with each other – that they occur as components of a single phenomenal field; to put the same point in different terminology, that you enjoy a single phenomenal state that subsumes them all (Bayne, 2010, p. 75).

Here is an example of this single synchronic state, given by Bayne himself:

> I'm sitting in the Café Cubana (47 Rue Vavin, Paris). I have auditory experiences of various kinds, I can hear the bartender making a mojito; I can hear the dog behind me chasing his tail; and there's a rumba song playing somewhere on a stereo [...] I am enjoying visual experiences of various kinds. I can see these words as they appear in my notebook; I can see the notebook itself; and I have a blurry visual impression of those parts of the room that lie behind the notebook. Co-mingled with these auditory and visual experiences are olfactory experiences of various kinds (Bayne, 2010, p. 5).

Now Chadha concedes that there is *prima facie* phenomenological evidence for the existence of a single unified phenomenal state, the parts of which are all phenomenal experiences at any given time. But she argues that this evidence is defeated on closer examination. Here is what she writes:

> The unified conscious state he [Bayne] is in includes manifest experiences that are at the forefront of his awareness, and a host of other peripheral conscious experiences, which can perhaps be thought of as background states [...]. Given that there are many parts of any given conscious state, and we are not fully aware of all the parts, or not aware of all the parts to the same degree, it is hard to make sense of these states as part of the total conscious state at any given moment of time (Chadha, 2023, pp. 100-101).

There are different degrees of consciousness: not all conscious states are in the focus of attention, says Chadha. From this, she concludes that there is no composite, unified phenomenological field. But why should this be?

The very thought of a unified field with a centre and a periphery (the different degrees of consciousness) seems very straightforward. Bayne's description of the total experience in the Café Cubana – a description also reported by Chadha – is precisely that of a unified phenomenological field with different intensities. And not only is this unified field not difficult to think of, but it is rather easy to describe and 'understand', in the sense that everyone understands the meaning of Bayne's description.

So, the *prima facie* phenomenological evidence for the unity of consciousness is not threatened by Chadha's considerations. But then: if it can be said that there is phenomenological evidence for the existence of MMPs, then the same can be said for the existence of a total phenomenal field (the two evidences, if they are such, are of equal strength). Furthermore, the body as a composite entity also seems to be phenomenologically given (this evidence also has the same strength as the previous two). And the unity of the

mind-consciousness with the body also seems phenomenologically given.[11] Finally, the apparent mind-body unity is apparently a synchro-diachronic one. So, if (the impression of) being phenomenologically given is a good reason for admitting MMPs, then one should also admit the synchro-diachronic unity of mind-body, i.e. admit the existence of ordinary persons. But if so, then the self exists (at least as an ordinary person) and persists in time.

To conclude: Chadha had to defend these two theses: 1) Yes, there are MMPs, a certain kind of wholes, and 2) No, there are no other wholes – in particular, there are no persons. But we have seen that by 'phenomenologically' admitting MMPs, she is also led to admit persons: 1) Yes, there are MMPs, and 2*) Yes, there are other wholes – in particular, there are (ordinary) persons.

Strategy 2. Indispensable explanatory role

The second strategy for admitting MMPs has to do with their indispensability as explanatory-causal factors: for Chadha, MMPs are indeed explanatorily indispensable.

A paradigmatic case she provides is the following: the viewer's annoyance at an 'out-of-sync' film can only be explained by the existence of a previous audio-visual MMP; the disappearance of this MMP is precisely the cause of the viewer's annoyance (Chadha, 2023, p. 107). But if MMPs have an indispensable explanatory role (IER), then they must be allowed to exist. So, yes, there are MMPs (a certain kind of whole), because they have an IER.

Then Chadha needs a "no": "No, there are no other wholes – in particular, there are no persons", and this is true, one should say, because other wholes, including persons, do not have an IER. Let us just assume that this is true.[12] So it seems that we have our desired yes and no answer: yes, there are MMPs, a certain kind of wholes, because they have an IER and no, there are no other wholes – in particular, there are no persons, because they do not have an IER.

But is that the case? I doubt it: I think Chadha is a bit overconfident about this thesis: "Yes, there are MMPs, because they have an IER". Let me try to explain why. Consider "Case 1" and "Case 2":

> Case 1
> For many months, a baby, like Pavlov's dog, has had the experience of a bell ringing when food is brought to it, until at time t it hears the bell but does not see the food, and this surprises it.
>
> Case 2
> When a certain colour is perceived, a certain smell is also perceived. This creates a habit that is 'contradicted' at time t, when the colour is perceived but not the smell, causing discomfort.

In these two cases, the reaction of surprise or discomfort is explained without reference to MMPs – MMPs do not appear in the explanation. Therefore, in order to explain the observed effect (surprise or discomfort), it is not necessary to assume that the two sensory experiences involved form an MMP. Let us now consider a third parallel case:

11 Even Descartes said something very similar: the difficult thing, he said, is to prove the mind-body distinction; their unity is a basic and primitive notion. See Shapiro (2007, pp. 63-65).

12 Merricks (2001), for example, offers some standard arguments for denying the existence of wholes on the grounds that they are causally/explanatorily redundant.

Case 3
When a film viewer perceives the word W, she also perceives the lip movement M. This creates a habit that is 'contradicted' at time t, when W is perceived in the absence of M, causing discomfort.

Just as in Cases 1 and 2 the reaction (surprise or discomfort) is explained without the need for an MMP, so, in Case 3, the discomfort caused by the out-of-sync phenomenon is explained without the need for an MMP: it is not true that the viewer's discomfort with an out-of-sync film can only be explained by the existence of a prior audio-visual MMP. Thus, it seems that one can avoid attributing an IER to the paradigmatic MMP considered by Chadha, i.e. the audio-visual MMP.

But what is true for a paradigm is, of course, also true for the other cases (otherwise, it would not be a true paradigm). That is, the conclusion generalises from the paradigm case to the rest of the MMPs: MMPs do not have indispensable causal/explanatory power. Thus, Chadha's reason for allowing MMPs fails.

But there is more: we have assumed that there are no wholes other than MMPs, and in particular that there are no persons, *because they have no IER.* So, if an alleged entity *x* has no IER, then *x* does not exist.[13] But if MMPs have no IER, then MMPs do not exist.

So: we had conceded that "No, there are no other wholes – in particular, there are no persons". But if MMPs do not have an indispensable explanatory role, then it is also true that "No, there are no MMPs".

The instability of Chadha's position

To sum up and conclude. Chadha wants to admit MMPs in order to answer the Nyāya philosophers' argument for the existence of the self. To do this, she needs to have a justification for this pair of propositions:

Yes, there are MMPs, a certain kind of whole.

No, there are no other wholes – in particular, there are no persons.

She offers two justifications for this pair of theses: the phenomenological strategy and the IER-based strategy. But the phenomenological strategy is too weak: it allows us to admit the existence of MMPs, but does not prevent us from admitting persons. The IER-based strategy, on the other hand, is too strong: it allows us to deny the existence of persons, but ends up denying the existence of MMPs.

So, Chadha's yes-and-no position (yes, there are MMPs; no, there are not persons or other wholes) is unstable: if it is justified in one way (via phenomenology), it ends up as a yes-yes position; if it is justified the other way (via IER), it becomes a no-no position.

5. Concluding summary

Vasubandhu, as we have seen, rejects any form of self/ego. Nyāya philosophers disagree: they argue that the self is necessary to explain MMPs. Chadha replies on behalf of Vasubandhu: MMPs exist, but this existence does not require a self. But this reply has two problems:
1) Chadha's thesis is incompatible with what Vasubandhu says about what exists, and
2) Chadha's position is unstable, given the strategies of justification she offers.

13 Indeed, according to the Vasubandhu read by Chadha – and by Gold (2015) – to exist is to have indispensable causal/explanatory power. Hence: no IER, no existence.

REFERENCES

Sanskrit Sources

AKBh = *Abhidharmakośabhāṣyam of Vasubandhu*. Ed. P. Pradhan & A. Haldar. Patna: K. P. Jayaswal Research Institute, 1975;

NSBh = *Gautamīyanyāyadarśana with Bhāṣya of Vātsyāyaya*. Ed. A. Thakur. New Delhi: Indian Council of Philosophical Research, 1997.

Other Works

Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press;

Chadha, M. (2023). *Selfless Minds. A Contemporary Perspective on Vasubandhu's Metaphysics.* Oxford: Oxford University Press;

Dasti, M. (2023). *Vātsyāyana's Commentary on the Nyāya-sūtra:* A *Guide.* Oxford: Oxford University Press;

Dasti, M., & Phillips, S. (2017). *The Nyāya-sūtra: Selection with Early Commentaries*. Indianapolis, IN: Hackett;

Gold, J. C. (2015). *Paving the Great Way. Vasubandhu's Unifying Buddhist Philosophy*. New York, NY: Columbia University Press;

Gold, J. C. (2022). Vasubandhu. *The Stanford Encyclopedia of Philosophy.* Eds. E. N. Zalta & U. Nodelman. Retrieved from https://plato.stanford.edu/archives/win2022/entries/vasubandhu;

Guerrero, L. (2023). Ontological pluralism in abhidharma debates about the existence of past and future dharmas. *Philosophy East and West*, 73(2), pp. 264-285;

Kapstein, M. (2001). *Reason's Traces*. Sommerville: Wisdom Publications;

Kellner, B., & Taber, J. (2014). Studies in Yogācāra-Vijñānavāda idealism I: The interpretation of Vasubandhu's *Viṃśikā*. *Asiatische Studien - Études Asiatiques*, 68(3), pp. 709-756;

Merricks, T. (2001). *Objects and Persons*. Oxford: Oxford University Press;

Shapiro, L. (Ed.) (2007). *The Correspondence Between Princess Elisabeth of Bohemia and René Descartes*. Chicago, IL: University of Chicago Press;

Siderits, M. (2021). *Buddhism as Philosophy*. Indianapolis, IN: Hackett;

Tomasetta, A. (2025), *Io e mondo nella filosofia indiana. Due tradizioni di pensiero oltre il canone occidentale*. Roma: Carocci;

Watson, A. (2017). Self or no-self? The ātman debate in classical Indian philosophy. In J. Tuske (Ed.), *Indian Epistemology and Metaphysics* (pp. 293-317). London, NewYork: Bloomsbury.

LUCA ZANETTI
University of Bologna
luca.zanetti10@unibo.it

I AM NOT THE BEARER OF EXPERIENCE[1]

abstract

It seems to me that I am the bearer or owner of experience. However, this impression is misleading. If I am external to experience (that is, if what I am is not reducible to the way in which I appear in the present experience), then I can't know with absolute certainty that I exist. But I do. Hence I am not external to experience, and a fortiori *I am not the external bearer of experience. If I am internal to experience, then what I am is reducible to the way in which I appear in the present experience. However, if I am a part of experience, I can't be its owner or bearer, for an owner or bearer of something can't be a part of what it bears or owns. Therefore, I am not the internal bearer of experience.*

keywords

self, self-consciousness, certainty, The Transcendence of the Ego, sense of ownership

1 An earlier draft of this article was presented at San Raffaele University, in October 2024, during the conference *Exploring Personal Identity. Philosophical Perspectives and Insights from Arts*. Special thanks are due to the organizers and to all those in attendance on this occasion. I am especially indebted to Alberto Barbieri, Andrea Bottani, Edoardo Fugali, Francesca Forlè, Sara Heinämaa, Alfredo Tomasetta, Elisabetta Sacchi, Federico Zilio for their feedback during the conference. The ideas articulated in this article take inspiration from the work of Buddhist teacher Franco Bertossa and the research team of the ASIA Research Center, Bologna, to whom I am deeply grateful.

Phenomenology and Mind, n. 29 - 2025, pp. 90-97
DOI: 10.17454/pam-2901
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

It seems to me that I am the *bearer* or *owner* of the present experience. However, in this paper I shall argue that this impression is misleading.

A bearer of experience can be internal or external to experience. Something *s* is internal, relative to some experience *e*, iff *s* is identical to its phenomenal character in *e*, that is, to the way in which it appears in *e*, to what it's like to experience *s* in *e*. Something is external, relative to some experience *e*, iff *s* is not internal, relative to *e*. To illustrate, my dog is external to experience (at least *prima facie*, unless solipsism is true, say), whereas a particular shade of blue is internal to experience (at least *prima facie*, unless some form of naive realism[1] is true, say).

The distinction between being internal or external to some experience is exhaustive and provides us with a way of classifying views about the nature of the self. The argument of this paper is that both views that regard the self as internal and views that regard the self as external face problems in justifying the claim that the self is the bearer or owner of experience.

This way of carving up the logical space of views on the nature of the self is not standard, so it needs some comments before proceeding.

First, something can be internal to some experience *e* if it is a *part* of *e*, but also if it is *identical* with *e*. In other words, something can be internal to *e* by being identical with some *sub-part* of the total phenomenal character of *e*, but it can also be internal to *e* by being identical with the *total* phenomenal character of *e*. Also, to be clear, there is no distinction between *e* and, as we say, *its* phenomenal character: the present experience *is* its phenomenal character.

Second, there can be self-experiences, that is, the self can have some phenomenal character, even if the self is *not* internal to experience. That is, in some views the self is not reducible to some phenomenal character, and yet the self can (or even must) appear in experience. Some forms of animalism regard the self as something that can be experienced even if it is external to experience, that is, even if it is something that is not reducible to the way in which it appears in experience.

Third, it is important to stress that the *extension* of the predicates 'internal' and 'external', as defined above, is itself a controversial philosophical matter. There might be nothing that is

1 For discussion of recent naive realist views see Genone, 2016.

external to experience (if solipsism is true, say), and there might be nothing that is internal to experience (if some form of naive realism is true of *all* that appears in experience).

Fourth, if something is internal to experience, then *a fortiori* it *can* be experienced. This notion – being internal to experience – is meant to be neutral among different views about self-consciousness. Something internal to experience can be experienced *pre-reflectively*, *reflectively*, or in any other ways. For something to be internal to experience is for it to be identical with the way in which it *shows up*, that is, with the way it *appears*. In other words, however we think about self-consciousness, for something to be internal to experience is for it to be identical with *some* phenomenal character.

Since this way of carving up the logical space is not standard in contemporary debates, some recent theories of the self might not have made fully explicit whether the self is reducible to the way in which it appears to the present experience. That being said, recent views that seem to regard the self as internal to experience include, among others, (Strawson, 2017), (Zahavi, 2014), (Ramm, 2017). To illustrate, in Strawson's view, every single experience is a self, for what I am now just is the totality of the present phenomenal character. In Ramm's view, the self is internal to experience because it is identical with a specific part of experience that is there in every experience, namely a phenomenal character that he describes, following (Harding, 1988/2001, p. 109), as a "visibly boundless Space".

Recent views that regard the Self as external to experience include, among others (Dainton 2008), (Olson, 1997), (Snowdon, 2014), (Hudson, 2001) and (Parfit, 2012). To illustrate, according to Snowdon and Olson, I am a biological organism, so I am a body, and a body is not not reducible to some phenomenal character.

Let us now look at the arguments for the conclusion that the self is not the bearer of experience. In §2 I shall argue that I am not external bearer of experience; in §3 I shall briefly mention and put on a side the view that there can be an external bearer of experience that is not me; in §4 I shall argue that I am not the internal bearer of experience.

2. I am not an external bearer of experience

The argument for the conclusion that I am not an external bearer of experience runs as follows. First, I can know non-inferentially and infallibly that I exist. Second, I can know non-inferentially and infallibly that I exist only if I am not external to experience. Therefore, I am not external to experience. And if I am not external to experience, then, *a fortiori*, I am not the *external* bearer of experience. So, I am not the external bearer of experience. Let's look at the motivation for each premise.

I can suppose that right now I am deceived by an (almost) all powerful evil demon who is creating the impression that some propositions are true even if in fact they are false. As I am so supposing, it seems that there are some truths of which I can be absolutely certain: that there is *this* (what I call "the present experience"), that I am thinking (*cogito*), that I exist (*sum*).

This is Descartes' point (see his *Reply to the Second Set of Objections,* Descartes, 1984, p. 100), at least on some interpretations of it (for a different reading see e.g. Peacocke, 2012). Here we are concerned with the truth that *I exist*. The first premise of the argument says that it is possible to have knowledge of this truth that is both *non-inferential* – it is based on the present experience only – and *infallible* – its ground is sufficient to rule out the possibility that I don't exist. For brevity's sake, I shall sometimes speak of *certainty* in order to refer to knowledge that is both non-inferential and infallible.

This premise can be challenged (and we shall see in a moment some ways to do so). But I take it that to reject this premise comes with some significant cost. It does *seem*, to say the least, that even in the most radical skeptical scenario I can still know with certainty that I exist. Any theory that rejects this premise needs to *explain away* its apparent truth.

Let's move to the second premise of the argument. The claim that I can know non-

inferentially and infallibly that I exist is compatible with the claim that I am external to experience *only if* it is also possible to know non-inferentially and infallibly the existence of something that is external to experience. However, in the sort of radical skeptical scenario that we envisage in a Cartesian-like meditation, we are supposing that all beliefs about things that go beyond the present experience might be false. In particular, it seems to be impossible to know non-inferentially – that is, on the basis of the present experience *alone* – and infallibly the existence of some particular thing that is not identical with (or a sub part of) the present experience. This claim is, if anything is, virtually universally accepted in contemporary debate in epistemology. Thus, the supposition that I am external to experience is not compatible with the claim that I can know non-inferentially and with certainty that I exist.

To clarify, let's suppose, *for reductio*, that I am identical with some particular body (and of course, let us also suppose, for the argument's sake, that a body is external to experience). The present experience is not sufficient to know with certainty that some particular body exists. I can have the experience *as of* my harms, *as of* my head, *as of* my feet, etc., but, given the supposition that I might be living in a radical skeptical scenario, the existence of the present experience does not guarantee that there actually exists a body there.

Consider, by analogy, this case. I am in a park and see what looks like a rabbit that is hidden behind the trees. In fact, I am only seeing its muzzle. For there to be a rabbit there must be something more than a muzzle. It might be that behind the trees there aren't any of the other parts that are regarded as necessary for something to be a rabbit. Hence, I can't know non-inferentially and infallibly that there is a rabbit there.

The same applies in the case of my own existence if I regard the self as external to experience. It seems that I exist (this impression is the same thing as seeing the muzzle), but I take it that for me to exist something more than the present experience must exist as well (the experience of the muzzle is not enough for concluding infallibly that the rabbit exists). Now, experience *alone* is not enough to conclude with certainty that what seems to exist out there really exists. Hence, if I am external to experience, then I can't know with certainty that I exist. But I do know non-inferentially and infallibly that I exist (or so the first premise says). Hence I am not external to experience.

Note that I am *not* claiming that this argument is by itself sufficient to establish the claim that the self and experience are *immaterial*, and anyway the aim of the argument is not to reach that conclusion[2]. The argument only aims to establish the conclusion that the self exists and is *internal* to experience – either identical with it, or with some sub-part of it. It does not aim to say anything about the *nature* of experience (and hence, of the self).

Let's now consider some ways to reject the first premise of the argument – which is, I think, the most controversial one. One way to reject the first premise is to deny the existence of the self. If there is no self then I can't know with certainty that I exist. We might have the impression of existing, but this impression is misleading. However, if I don't exist, then *a fortiori* I can't be the bearer of my experience, and this is enough to establish the point of this paper. If there is no self then in some important sense *no one* bears the present experience (though see the hypothesis mentioned in the next paragraph).

One might reject the claim that it is possible to know the existence of the self *non-inferentially*. Yet, it is very hard to reject this premise – if one is not rejecting it on the ground that there is no such a thing as the self. The premise is in fact quite humble: it says that it must be *possible* to know of one's own existence non-inferentially. Descartes and many others

2 See Swinburn (2018) for a recent discussion of Descartes' indubitability argument for the immateriality of the self.

after him claimed that it was possible[3]. So, again, to reject the possibility of non-inferential knowledge of one's existence will require, to say the least, that one explains away a quite strong and widely shared seeming.

It is perhaps easier to reject the claim that it is possible to know one's own existence *infallibly*. But even if we were to reject the possibility of infallible knowledge of one's own existence, still it seems true that we can have *fallible* knowledge of our own existence, and that this knowledge can be gained *non-inferentially*. Consider, again, the skeptical scenario in which the existence of the external world is a sheer illusion. Still, while I am so supposing, I am having an experience, and the experience seems to be the experience of *me* having it. It is *me* who is doubting, and in order to know that it is *me* who is doubting I do not need to appeal to any further principle that functions as a bridge from the existence of the present experience to my existence. Perhaps I *might* appeal to some such principle, but I *need not* to.

Even if we suppose that this experience does not provide me with *infallible* grounds for concluding that I exist, still I think that most of us will concede that it provides us with *strong* grounds indeed to so believe. And, arguably, the *best explanation* why the present experience is enough to provide strong grounds for concluding *non-inferentially* that I exist seems to be that the self is *internal* to experience.

So, even if one has some independent grounds for believing that the self is external to experience, some story must be told to explain why, even if the Self is external to experience, it can be known *non-inferentially* and with *infallible* or anyway *strong evidence* that the self exists. I do not deny that there might be some *prima facie* plausible story to be told here. I just wish to highlight that there is a challenge to theories that regard the self as external to experience: the challenge is that it is hard to explain how it is possible for me to know non-inferentially and infallibly (or, at any rate, with strong evidence) that I exist if I am not something internal to experience.

3. Can there be an external bearer of experience that is not me?

Suppose that it is true that I am not the external bearer of experience. Does this rule out the possibility that there is an external bearer of experience that is not me? One might think, as Frege at some point speculates in his *Der Gedanke*, that the bearer of experience is some divine transpersonal entity. He writes:

> It is so much of the essence of each of my ideas to be the content of my consciousness, that every idea of another person is, just as such, distinct from mine. But might it not be possible that my ideas, the entire content of my consciousness might be at the same time the content of a more embracing, perhaps divine, consciousness? Only if I were myself part of the divine consciousness. But then would they really be my ideas, would I be their bearer? This oversteps the limits of human understanding to such an extent that one must leave its possibility out of account. (Frege, 1956, p. 300)

Another view might be that there is something that is not me but *somehow* belongs to me – perhaps a body – which is the owner or bearer of the present experience. In these views, I exist, there is a bearer of what I ordinarily regard as my experiences, but this bearer is not me. Although I do not wish to argue here that these views are false, it is important to stress that the very idea of a bearer of what appears as my experience that is not me is, to say the least,

3 A classic reference on this point is of course Husserl (1931/1960). For a recent defence of this claim – and, more generally, of the argument from certainty articulated in this section – see Duncan (2015) and the literature referred to therein.

prima facie puzzling. However, the puzzlingness of the view in part depends on the apparently intuitive thought that the self, if it exists, is the bearer of experience. But if this view turns out to be false – as I am arguing in this paper – then perhaps we can make room in our way of conceptualizing experience for the idea of something that is not *me* and yet is the bearer of what nonetheless prima facie appears as *my* experience. This could be a way to make room for the idea of a transpersonal bearer of experiences, or even, as Frege speculates in the passage quoted above, for some divine consciousness. To see why this can make any sense at all we need, however, to make some steps forward in the argument.

4. I am not the internal bearer of experience

I have excluded the possibility that I am the external bearer of experience on the ground that I can know with certainty that I exist and that this knowledge is possible only if I am internal to experience. The previous argument from certainty is a direct argument for believing that the self exists and is internal to experience. If we have some independent grounds to think there is a bearer of my experience, and that this bearer should be me, then with the argument from certainty we can conclude that the bearer of experience is internal. However the view that the bearer of experience is internal faces its own problems.

Consider, first, the view according to which the self is identical with the total phenomenal character of a given experience. In this view, it is hard to understand how the self can be said to *bear* or *own* experience. For, in this view, the self just *is* the present experience. What kind of *ownership* can experience have with itself – beside, precisely, *being identical with itself*?

One way to articulate the reason why self-identity can't be self-ownership is the following. It seems that something can bear or own something only if there is some *distinction* between that which is own and that which owns. Everyday ascriptions of ownership always seem to concern distinct objects: I own my bike, I bear my thoughts, I own my rights, I own my body etc. True, sometimes we can say things like "I own myself" or "own yourself", but these are just ways of specifying relations between oneself and one's own thoughts, desires, choices, etc. so these cases do not count as counterexamples.

Consider now the view according to which the self is identical with a particular *part* of the total phenomenal character of the present experience. This view does not face the problem of identifying the self with that which it is supposed to own. However, this view faces other problems.

First of all, how can a part of some larger whole be also the bearer or owner of the whole? It seems that something can bear or own something else not only if there is a difference between the two, but also only if the former is not a part of the latter. Everyday ascriptions of ownership conform to this observation: I own my bike, but I am not part of the bike, I possess my body but I am not part of my body, etc. No star, however shining, can own the sky, just as no self, however sui generis its phenomenal character might be, can own the present experience.

Second, one can perhaps try to save the claim that the self owns its experiences by saying that the self, a sub-part of the total present experience, owns the other sub-part of the present experience. Thus, to illustrate, if a proper characterization of the present experience features sounds and a self, then one can perhaps say that the experience of sounds belongs to the self, that is, to me. But even if the self-part of experience is *sui generis* in its phenomenological status, still, from a metaphysical point of view, *it is on a par* with all the other things that are internal to experience. After all, the self is just a phenomenal character like any other thing that is reducible to a phenomenal character (e.g., colors, sounds, etc., depending on one's theory). So, again, in what relevant sense of ownership is a self, conceived of as a part of experience, can bear the other part of experience?

Similar insights are elaborated in Sartre's discussion of the Ego in his *The Transcendence*

of the Ego. Sartre argues, as I just did, that if the Ego is internal to consciousness, then it can't be said to be the owner of consciousness. What he calls "consciousness" – what I have called "experience" – has a priority over the Ego, for the Ego itself appears within consciousness. Here is one passage in which he makes the point:

> As the me is an object, it is obvious that I will never be able to say: my consciousness, i.e. the consciousness of my me (except in a purely designating sense, in the sense in which one says for example 'The day of my baptism'). The Ego is not the proprietor of consciousness, it is its object. (Sartre, 2004, p. 26)

And few lines later he says:

> I can thus formulate my thesis: transcendental consciousness is an impersonal spontaneity. (Sartre, 2004, p. 27)

Sartre's claim in this book is that the Ego is *transcendent*, it is an *object* within consciousness, and he relies on *this* claim in order to conclude that the Ego cannot be the owner of consciousness. However, there is no need to claim that the Ego or the self is an object in order to reach this conclusion. In fact, my argument against the view of the self as the internal bearer of experience does *not* rest on the premise that the self is experienced as an *object*. The argument applies to *any* view that regards the self as internal to experience, that is, as identical with the way in which it appears in experience. It doesn't matter *how* it appears – whether it appears as an object or as a subject, reflectively or pre-reflectively, whether it is identical with the whole present experience or just to some part of it. In other words, it doesn't matter what the correct account of its *phenomenal character* is. All it matters is that it *is* a phenomenal character. For, if the self appears in experience, then it is part of experience or identical with it, and as such cannot count as the bearer of experience, in any relevant sense of the term. This is why there is an important sense in which experience – what Sartre calls "transcendental consciousness" – is *impersonal*. If the self exists, and it is internal, then there is the appearance of a personality, but this appearance itself – and the totality of experience – is impersonal. There is the experience of a self, but precisely because there is this experience – and a self is nothing more than what is given in experience – the self is not its bearer.

5. Conclusions

I have presented arguments that challenge the claim that I am the bearer of experience. If these arguments are on the right track, then we should reconsider the way in which we pre-analytically think of the self and of its relationship with experience[4]. We think of the self as the *subject* of experience, and the subject is also often conceived of as the *owner* or *bearer* of experience (but see Albahari, 2006 for discussion of this widespread assumption). But perhaps we should carefully distinguish the two notions – that of a *subject* of experience, and that of a *bearer* of experience – and say that I can be the subject of experience without being its bearer. To be the subject of experience is for there to be a *phenomenal character* in experience that justifies the use of that notion. To speak of me as the subject of experience is then to make a *phenomenological* observation, and perhaps a true one. But to speak of me as the *bearer* or *owner* of experience is to make a *metaphysical* claim about the relationship between me – the

4 From the claim that the self is not the bearer of experience one can also argue to the conclusion that the self does not possess free will. See Zanetti, 2022 for a Sartrean-inspired argument to that effect.

subject of experience – and experience. I appear in experience as its subject – and this is the phenomenological point – but I am not its owner or bearer – and this is the metaphysical point – and I am not its bearer or owner *precisely because I appear in it*. Or, if one prefers to claim that being the bearer of experience is necessary for something to count as a self and as a subject of experience, then the upshot of the argument might be that although I do *appear as* the self, the subject of experience and as its bearer, this *appearance* is false. Either way, I must make room in my way of thinking of what I ordinarily call "my" experiences for the thought that I do not own them. If I am internal to experience, then, there is a sense in which I have no privileged place in experience. Experience owns me. *Being a phenomenal character*, I am on a par with colors, sounds, thoughts, and anything that appears in experience.

REFERENCES

Albahari, M. (2006). *Analytical Buddhism: The Two-Tiered Illusion of Self*. London: Palgrave-Macmillan;

Dainton, B. (2008). *The Phenomenal Self*. Oxford: Oxford University Press;

Descartes, R. (1984). *The Philosophical Writings of Descartes, Volume 2*. Cambridge: Cambridge University Press;

Duncan, M. (2015). We are acquainted with ourselves. *Philosophical Studies*, 172 (9): 2531-2549;

Frege, G. (1956). The thought: A logical inquiry. *Mind*, 65 (259): 289-311;

Genone, J. (2016). Recent Work on Naive Realism. *American Philosophical Quarterly* 53 (1);

Harding, D. E. (1988/2001). *The Little Book of Life and Death*. London: The Sholland Trust;

Hudson, H. (2001). *A Materialist Metaphysics of the Human Person*. Ithaca, NY: Cornell University Press;

Husserl, E. (1931/1960). *Cartesian Meditations*. Trans. D. Cairns. The Hague: Martinus Nijhoff;

Nida-Rümelin, M. (2017). Self-Awareness. *Review of Philosophy and Psychology*, 8 (1): 55-82;

Olson, E. T. (1997). *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press;

Parfit, D. (2012). We are not human beings. *Philosophy*, 87 (01), pp. 5-28;

Peacocke, C. (2012). Descartes Defended. *Aristotelian Society Supplementary Volume*, 86 (1): 109-125;

Ramm, B. (2017). Self-Experience. *Journal of Consciousness Studies*, 24 (11-12): 142-166;

Sartre, J.-P. (2004). *The Transcendence of the Ego: A Sketch for a Phenomenological Description*. London: Routledge;

Snowdon, P. F. (2014). *Persons, Animals, Ourselves*. Oxford: Oxford University Press;

Strawson, G. (2017). *The subject of experience*. New York, NY: Oxford University Press;

Swinburne, R. (2018). Cartesian Substance Dualism. In J. Loose, A. Menuge, & J. P. Moreland (Eds.), *The Blackwell Companion to Substance Dualism* (pp. 133-152). Oxford: Wiley-Blackwell.

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.

Zanetti, L. (2022). Un argomento fenomenologico contro l'esistenza del libero arbitrio. *Verifiche*, anno LI, n. 1-2: 277-301.

FEDERICO ZILIO
Department of Philosophy, Sociology, Education and Applied Psychology – FISPPA, University of Padua, Italy
federico.zilio@unipd.it

METAPHYSICAL ACCOUNTS OF PERSONHOOD AND THEIR ETHICAL IMPLICATIONS FOR THE VEGETATIVE STATE: A COMPARATIVE ANALYSIS

abstract

This paper investigates how metaphysical theories of personhood carry significant ethical implications regarding attitudes towards patients with disorders of consciousness. After providing conceptual clarifications on key terms such as moral status, vegetative state, metaphysical and moral personhood, and sortals, the paper analyzes five main approaches to theories of personhood: personism, animalism, the disjunctive/hybrid view, the constitution view, and ontological personalism. Related bioethical implications are presented for each theory: the intermittent person (personism, constitution view); personhood as accidental (animalism, disjunctivism); the indefinite substratum (disjunctivism); the dual human-person subject (constitution view); the non-rational person (ontological personalism). Ontological personalism appears preferable in dealing with liminal cases, anchoring criteria for identity in a corporeal entity of a rational nature, while recognizing radical changes in the manifestation of psychological characteristics. From a pragmatic perspective addressing uncertainties of vegetative state diagnoses.

keywords

person, vegetative state, sortal concepts, metaphysics, neuroethics

Phenomenology and Mind, n. 29 - 2025, pp. 98-110
DOI: 10.17454/pam-2909
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

This paper discusses whether and how metaphysical theories of the human person can also have moral implications, particularly on attitudes and obligations towards patients with disorders of consciousness, in particular patients in a chronic vegetative state.

Several philosophical questions about the person are related but different, such as the question of the metaphysical essence of the person (what it is to be a person, as opposed to a nonperson), the problem of personal identity (spatio-temporal permanence – persistence question – and the conditions for reidentification – evidence question), and the problem of the person in a moral sense (what moral implications for an entity that is recognized as a person) (Olson, 2007). In this regard, it is important to recognize the distinction between metaphysics and ethics of the person and avoid deriving a complete set of moral principles solely from metaphysical theories. Doing so risks committing the is-ought fallacy, as descriptive statements cannot logically entail prescriptive claims.

However, while metaphysical commitments cannot logically entail moral conclusions, certain theoretical approaches may carry prescriptive force by informing, influencing and constraining ethical considerations. Specifically, the connection between metaphysical and moral issues seems particularly strong when considering questions of personal identity and bioethics (or neuroethics). It seems reasonable to think that criteria for the existence and persistence of personhood could shape moral attitudes concerning moral personhood, in particular towards individuals in blurry and complex situations, such as patients with disorders of consciousness. I will analyze some metaphysical theories of the person, discussing potential ethical implications for the care and management of patients in vegetative states, a specific condition under the broader category of disorders of consciousness.

1.1 Conceptual Clarification and Some Premises

First, some definitions are necessary to avoid misconceptions:

- Moral status: *X* has moral status if and only if *x* is worthy of direct moral consideration. By direct moral consideration I mean the identification of non-derivative and non-mediated obligations or values, as the object of evaluation is the entity itself by virtue of its intrinsic values.[1] By indirect moral consideration, on

1 Moral status can be conceptualized as either a graduated spectrum or a binary "all or nothing" structure (DeGrazia, 2008). The graduated view allows entities to have different degrees of moral status, meaning that moral agents

the other hand, I mean the identification of derived and mediated obligations, i.e. those which, although addressed to that entity, derive their motivation from some relationship with other entities.

- Moral personhood: If an entity is regarded as a person, then it has full moral status (i.e., highest degree) and vice versa (Jaworska & Tannenbaum, 2019). In this sense, we can speak of the dignity of the person, which produces an effect of recognition respect (Darwall, 1977) that takes the form of stringent duties of non-interference, non-maleficence, and fair treatment, as well as – not as stringent – demands for assistance and of beneficence.
- Vegetative state: The vegetative state (VS)[2] is a condition due to severe brain injury in which the patient is in a state of wakefulness and maintains neurovegetative functions (e.g., autonomous breathing, sleep-wake cycle and spontaneous eye-opening, digestion, metabolism, thermoregulation), but without any clinical symptoms of self-awareness, awareness of the environment, and preserved cognitive functions (the patient may show some non-purposeful reflexive motor behavior and facial expressions) (Schnakers & O'Brien, 2023).

In light of the frequent possibility of recovery of consciousness even after several months or years, various medical associations and institutions have recently begun to reject terms such as "permanent" and "irreversible" vegetative state, replacing them with "chronic" or "prolonged" (Giacino et al., 2018; Kondziella et al., 2020). Moreover, there is a large body of empirical evidence showing that there is still a significant rate of misdiagnosis of the vegetative state (i.e., cases of covert consciousness and cognitive-motor dissociation) (Young et al., 2021). However, for the sake of this specific theoretical argument analysis of theories of the person, when referring to the vegetative state, I will assume an ideally 100% certain diagnosis, both in terms of the absence of awareness and the absolute impossibility of future recovery.

Second, to circumscribe the target of the paper, some other premises are necessary to determine the range and limits of the argument:

- I will refer to issues related to human beings, although some of the theories of personhood here analyzed may also refer to nonhuman entities.
- I will refer to real-world non-branching cases, specifically the vegetative state.
- I will only refer to accounts that support that personhood is derived from properties related to human beings as such. I will not consider those theories that understand personhood (ontologically and/or morally) as an extrinsic product of special relations or conventional acts.

2. Personhood and Vegetative State

There is a sort of misunderstanding about the concept of person that we use in everyday language, a misunderstanding that has reverberated in the bioethical domain and more recently extended into the field of neuroethics (Blumenthal-Barby, 2024). This issue becomes clear if we take the vegetative state as a case in point. Consider the following two statements:

have obligations towards all entities with moral status, but with differing weights depending on the situation and competing values. Instead, the binary view holds that entities either have moral status (with no degrees) or none at all. For the purposes of this work, a gradualist conception of moral status is adopted, as it is better equipped to provide adequate nuanced responses to the complex challenges of contemporary bioethics and neuroethics.

2 In 2010, the alternative notion of Unresponsive Wakefulness Syndrome (UWS) has been proposed (Laureys et al., 2010).

a) "A human being in a chronic vegetative state is not a person anymore and therefore we have no obligations towards them as such.".
b) "A human being in a chronic vegetative state is still a person and should be respected as such".

First, these two statements are not opposed with respect to the moral meaning to be given to the person. Both could support the idea that the person has as its intrinsic dignity; also, both sentences make sense to us. Indeed, the VS patient is no longer conscious, i.e., aware neither of himself nor of the world around him, so it may seem plausible to think that the person – somehow – is no longer there; on the other hand, that bodily individual who still breathes autonomously, has sleep-wake cycles, opens his eyes, and produces some reflex movement seems a human person like us, even if he is not conscious anymore. This represents indeed a challenging boundary case regarding the ascription of moral status. The metaphysical question of whether the VS patient is still numerically identical to their previous self, or now a distinct entity, carries significant ethical weight. How different metaphysical theories characterize personal identity and criteria for persistence can influence the moral obligations we have towards VS patients.

The difference between the two sentences lies in the metaphysics of the person. Indeed, these two statements are based on different criteria of existence and persistence of the person, entailing not only theoretical but also moral consequences for personhood. The first statement is based on the functionalist account of the person, while the second on the substantialist one. The neo-Lockean functional view argues that the person is an intelligent entity possessing a set of mental functions or properties that allows the continuity of the self over time. The substantialist neo-Thomist view argues that the person is an individual entity that exists according to a rational nature, which also includes the vegetative and sensory faculties. From these two conceptual paradigms, a series of contemporary theories of the person emerge, which I will analyze and compare.

3. A sortal approach to "person" and "human being"

It is possible to categorize some approaches to personhood in the field of bioethics and neuroethics of DOCs. These categories are characterized by the diverse ways in which the concept of person is understood from a metaphysical perspective.

One way to distinguish these theories is to use the concept of sortal. "Sortal" is a term of Lockean origin but actually refers to the Aristotelian distinction between substance and accidents. Following David Wiggins' theory, sortal concepts serve to provide the principle of individuation of continuants, that is, entities that continue to exist (Wiggins, 2001). According to Wiggins, in fact, a statement of individuation needs not only a spatiotemporal continuity of the entity, but also a concept capable of pointing to that entity. To understand the identity of an object, it is essential to link it to the sortal under which the object falls; in this sense, identity is dependent on the sortal concept, understood as that concept that answers the Aristotelian substantive question "what is it?" and provides principles of individuation to the object it instantiates. So, to say that a particular entity is the same as another in a particular time is to say that there is a general term or sortal that specifies the criteria of identity, persistence, and existence by providing the relation between different occurrences of that object.

Wiggins differentiates substance sortals from phase sortals. While substance sortals answer the question "what is x?" or "what sort of substance is it?" (identification and reidentification), which applies to an entity at every moment of its existence ("a dog", "a tree", etc.), phasal sortals are a specific type of restriction of the substantive ones and apply to an entity but only when this entity has certain properties ("tall", "big", "short", "brown", "young", "old", "seedling", "puppy", etc.); in other words, it answers the question "what is/was/will be *x* at a certain period *t* of its entire existence?".

Now, through the lens of Wiggins' sortal theory, we can pose a series of questions: What are the criteria for identifying and tracing the persistence of the entity determined by the concept of personhood? Is personhood itself a sortal concept? If so, is it a substance sortal or a phase sortal? Can the person provide principles for individuating a definite, persistently existing entity, or is it merely a temporary phase of a more general entity? More specifically, the key issue for disorders of consciousness is: what becomes of personhood in such conditions? Is a patient in a vegetative state (VS) still a person? Or have they become some other kind of entity instead? For instance, are they now simply a human being who has ceased being a person? Or do they remain a person in all respects despite their condition?

The concepts of numerical identity and qualitative identity might be helpful for understanding the relationship between phase sortals and substance sortals in this context. "Numerical identity" is the relation that every object has only with itself and with no other and concerns the uniqueness and spatio-temporal continuity of an entity: x and y are numerically identical if and only if y=x.[3] For example, if one has a favorite mug that chips and discolors over time, it remains numerically identical despite those qualitative changes. Qualitative identity, on the other hand, concerns the similarity or sharing of properties between two objects. Two objects can be qualitatively identical, while being numerically distinct. For instance, two mugs from the same manufacturer may be qualitatively identical in shape, color, and pattern, yet they are still numerically distinct objects.

In the specific context of this paper, the questions concerning the concept of personhood pertain primarily to numerical identity. i.e., whether and when a person ceases to be numerically identical to themselves over time following huge qualitative, intrinsic changes. The sortal analysis aims to clarify the metaphysical criteria for the numerical identity, persistence and reidentification of personhood. Taken together, one can assume the existence of four sortal relations between person and human being:

1) the person as a substance sortal and the human being as a phase sortal;
2) the person as a phase sortal and the human being as a substance sortal;
3) the person and the human being as phase sortals;
4) the person and the human being as substance sortals.

Following this outline, I will briefly present the following theories: Personism (1), animalism (2), disjunctivism or hybrid view (3), constitution view and ontological personalism (4).

3.1 Personism

Personism[4] argues that the concept of a person refers to an entity possessing a range of psychological functions or higher-order cognitive processes, e.g., self-awareness, sentience, interests, rationality, capacity for intersubjective relations, agency, and moral autonomy (Levy & Savulescu, 2009; Mcmahan, 2010; Singer, 2011).[5] The person is a substance sortal in the sense of a being characterized by some of these capacities (depending on each sub-theory) and the

3 Numerical identity is grounded on the principle of the indiscernibility of identicals, which states that if two things are truly identical, then they share all their properties. From this follows that numerical identity therefore requires perfect qualitative identity.

4 I adopt the term "personism" from J. Teichman (1996), in order to avoid possible confusion with twentieth-century personalism as well as with ontological personalism (see below).

5 One of the early and influential definitions of personhood in bioethics that established the standard personist position was proposed by Michael Tooley (1972). Tooley argued that for an entity to be considered a person (and therefore have a serious moral right to life), it must possess self-consciousness as the ability to conceive of oneself as a distinct, persisting entity undergoing experiences (Tooley, 1972).

psychological connection over time of the self. In this view, consciousness (and its correlated neural processes) plays a crucial role as the basis and sustaining function of mental processes (psychological – mental – or embodied – neurological – accounts of identity). Therefore, the criterion of consciousness is seen as necessary and sufficient for personhood, while entities that do not yet possess or have permanently lost this characteristic are not considered persons. The human being, on the other hand, is viewed as a phase sortal in relation to the person, because the person could potentially be transferred to another human being or an artificial system. Regarding the vegetative state, since personhood is the substance sortal, numerical identity is based on existing as a person, not merely as a human living being: when personhood ceases, one ceases to exist in an essential way. Thus, the person never actually enters the VS, so that the patient is something numerically different from the preceding person.

3.2 Animalism

Animalism argues that we are human animals (members of the species *Homo sapiens*). Also, most animalists hold that human animals persist over time by virtue of brute-physical continuity (the continuation of biological life) rather than psychological continuity; therefore, psychological continuity is neither necessary nor sufficient for a human animal to persist over time. According to authors such as Eric Olson, what really identifies us is that we are human individuals who go through various stages of development, including that of being persons (Olson, 1997, 2007).[6] As human individuals with reference to the substance sortal "human animal" (or animal of some kind), our numerical identity begins as embryos and ends with the cessation of the functions of the brainstem that guarantee the integration of the entire organic system. We can therefore transition through phases like childhood, adolescence, adulthood, and old age (all of which are phase sortals) while remaining the same human animals underneath. Similarly, as human animals, we can acquire the status of persons (defined according to the functionalist approach) and then lose it, like a phase sortal. According to animalism, the VS patient is alive and the same numerical individual, because, despite cortical damage, brain areas such as the brainstem, thalamus, cerebellum, etc. remain preserved, allowing for the normal continuation of various neurovegetative functions (autonomous breathing, sleep-wake cycle, digestion, metabolism, thermoregulation, reflex movements, etc.). Thus, the individual still exists and is the same numerical entity that was previously conscious before entering the vegetative state and losing personhood as a phase sortal.

3.3 Disjuctive Approach/Hybrid View

The disjunctivist approach or hybrid view lies somewhere between personism, which holds that we are persons understood as psychologically persistent entities, and animalism, which holds that we are animals that persist through biological continuity (e.g., Langford, 2014; Noonan, 2021). This approach argues that we are – necessarily – purely psychological persistence nor purely biological ones, meaning that we are neither solely animal organisms nor solely psychological persons. Rather, we are bio-psycho-continuers, entities that can maintain numerical identity through either psychological or biological continuity, which are treated as non-mutually exclusive sufficient conditions (both being phase sortals).

6 For some animalists, the substantial sortal may not be so much the "human being", but rather the broader concept of an "animal of some kind". As David DeGrazia discusses, an organism could remain numerically identical even through radical changes like genetic modifications across species boundaries (e.g., becoming a member of a new hominid species). As he notes, «[w]hat's crucial is that we are essentially animals of some kind; we could not transform into plants or tables» (DeGrazia, 2005, p. 49). (I thank an anonymous reviewer for pointing this out).

The substance sortal is therefore the neutral continuer, while the person is a phase sortal connected with the psychological continuer, conceived in this case too through the functionalist psychological approach as an entity possessing a set of mental functions or properties that allows the continuity of the self over time. Therefore, the VS patient is numerically the same neutral continuer as before the brain injury, specifically a continuer who has permanently lost the psychological continuity condition but keeps existing as a continuer by virtue of their ongoing biological persistence.[7]

3.5 Constitution View

The constitution view is based on the idea that it is possible for two objects of different kinds, with distinct persistence conditions, to exist in the same space and time. As Lynne Rudder Baker (2000) argues, constitution is a unity relation very close to identity, such that if x constitutes y at a given time, then x and y do not exist separately, although they are not the same entity. According to this view, the human body and person belong to two different primary kinds (two substance sortals) and therefore are ontologically distinct, though united by the relation of constitution. Specifically, a person is constituted by a human body but is not identical to that constituent body (Baker, 2000). The relationship of constitution is the same as the relationship between the statue and the piece of clay that constitutes it: just as the statue is not identical to the piece of clay, so the human person is constituted by the human body but not identical to it. Thus, persons are constituted by human bodies but are not identical to them since the former essentially possess the capacity for first-person perspective (while the human body does not). According to the constitution view, since the person is what essentially defines us or what we refer to as a human individual, then in a vegetative state, only the human body exists, as the capacity for first-person perspective is lost (aligning with the functionalist approach insofar as first-person perspective requires consciousness).

3.6 Ontological Personalism

Ontological personalism argues that the substance sortal that identifies us is the concept of "human persons", which unifies both "human being" and "person". In fact, when referring to a "human being", we are not simply denoting a generic rational animality or mere biological membership in the species *Homo sapiens*, but rather precisely a (singular) human person.[8] The human person is defined as an individual corporeal entity that exists by virtue of a rational nature. The rational faculty, which also encompasses the sensitive and vegetative faculties, provides the vital principle of the living, sentient, and thinking entity. Therefore, the human person, in both animality and personhood, exists essentially as long as the vital faculties persist. In this view, a patient with vegetative state (VS) remains a fully human person, as the rational nature continues to inform the living body through vegetative functions (Eberl, 2020). In other words, the patient is still a person by virtue of their rational nature, even if unable to exhibit or exercise specific functions or properties (e.g., consciousness, cognition, language, autonomy, etc.). This departs from other positions:

7 In principle, according to the hybrid view, one could also conceive a case where the same individual continues to exist as a psycho-continuer without a biological continuer, for instance in a futuristic mind-uploading scenario.

8 According to the substantialist approach, the term "person" denotes an individual entity existing by virtue of a rational nature. Theoretically speaking, this could encompass entities other than human beings, such as certain non-human animals, future artificial systems exhibiting genuine rational capacities (not merely simulated intelligence), as well as angels and God itself. For this reason, ontological personalism specifies the more precise substantial sortal of the "human person".

the patient is neither a human animal who has lost personhood, nor a human being entirely disconnected from the identity of their previously conscious self. Rather, we still identify the patient as the same human person known before the pathological state, while recognizing – precisely because of the persistence of the same person – a radical difference in the manifestation of their rational nature – no longer as an actively exercised capacity, but as an inherent endowment (Kavanaugh, 2001).

4. Metaphysical Objection and Bioethical Implications

As mentioned in the Introduction, it is not possible to directly derive ethical consequences from the metaphysical level alone. Even if a theory might produce inconsistent outcomes from an ethical perspective, a robust moral theory of the person can still provide adequate duties, motivations, and recommendations for ethical action towards VS patients. Likewise, if a theory were to consider individuals in VS as persons in all respects, this would not automatically entail imposing absolute obligations such as protecting life at all costs or the inability to differentiate resource allocation between vegetative patients and those with other pathologies.

Nevertheless, certain metaphysical theories of personhood may be better or worse equipped to plausibly account for liminal cases like the vegetative state, avoiding inconsistent arguments, highly counterintuitive results, or objections regarding their explanatory power and comprehensiveness. On a metaphysical level, I argue that ontological personalism seems better able to account for certain issues inherent to the vegetative state compared to other theories, avoiding a number of problems, such as those that follow.

4.1 Metaphysical Objections

The intermittent person (personism, constitution view): If the sortal substance "person" depends on the contingent capacities of consciousness and cognition, a person's life could be highly intermittent, which arguably is, at least, a counterintuitive conclusion and conflicts with the usual meaning of the concept (e.g., a person "emerges" after the initial mental development of the young human being, but then "ceases to exist" following an accident, only to "return" upon awakening from a subsequent coma and a few months in VS, eventually "vanishing" gradually after a period of progressive dementia).

Personhood as accidental (animalism, disjunctivism): Treating personhood as a mere phase sortal (similar to "being a student") seems to contradict what we typically intend by the concept of a person, which connotes essential characteristics and the persistency of the entity we consider to be such. This may be accepted if we definitively separate the metaphysical level from the moral level of personhood (but in practice, this is unlikely to occur).

The indefinite substratum (disjunctivism): Referring to a neutral continuer as the substance sortal, with biological and psychological attributes as phase sortals, does not seem to offer sufficient criteria for individuation and persistence as required (what are we, precisely and fundamentally, not just accidentally?), nor practical factors for reference (we do not typically denote an indefinite, neutral continuer when speaking of a particular individual).

The dual human-person subject (constitution view): It is difficult to understand how the two primary kinds (person and human being) relate to each other in coexistence; in particular, how and why only the person possesses mental properties, despite sharing the same physical and neurophysiological basis for first-person perspective with the human being. Positing that both have mental properties seems to double the number of corresponding entities and subjects, while arguing that the human being has mental properties only derivatively from the person (Baker, 2000) may commit to a psychologistic position (Olson, 2007).

The non-rational person (ontological personalism): One could argue that in the vegetative state patient, rationality as general capacity for cognitive activity is lost, since the cortical

areas associated with such processes are compromised. Therefore, the patient cannot be considered a person. However, this objection can be addressed by focusing on the concept of rational nature. While functionalist theories are grounded on criteria like consciousness, cognition, sentience etc. as capacities and/or properties in execution, ontological personalism considers these as manifestations of rationality as a natural disposition intrinsic to the human being, regardless of developmental or operating level. In other words, VS patients do not become humans or animals of a non-rational nature, nor do they cease to be persons simply because they are unable to perform rational mental activities in a given time.

Taken together, ontological personalism seems preferable to other views for conceptual reasons, such as avoiding counterfactual implications about the cessation of existence due to temporary psychological changes, providing more stable criteria for identity and reidentification by anchoring them in a corporeal entity of rational nature, and being better equipped to handle boundary cases without denying personhood status based on mentalistic criteria.

4.2 Bioethical implications

As for the ethical implications of these theories, we can infer potential normative consequences for patients in vegetative state. Personism, animalism, disjunctivism, and the constitution view are based on or accept the neo-Lockean functional concept of person. However, the way the person is considered for each theory, as substantial sortal or phased sortal, has different ethical implications (leaving aside the possible addition of extrinsic ethical justifications and consequently indirect moral consideration).

Following personism, patients with VS do not have an inherent right to life or moral status as they are not persons anymore. The metaphysical dualism between "human being" and "person" can have significant ethical consequences. Since the human being is not numerically identical with the person who existed before the vegetative state, not only does patients have no serious right to life, but, in principle, since they have no moral status, there are no direct obligations towards them. Since the holder of rights and moral worth is only the person, there is nothing to prevent the patient in a vegetative state from being killed or kept alive for other purposes, such as organ donation with a beating heart in the absence of previously obtained valid consent, or, one could imagine, for experimentation with new potentially dangerous drugs, or hypothetical organ cultivation for use by other patients (a sort of "factory" for the continuous production of the liver in times of resource scarcity).[9] Not all proponents of personism may endorse such strong implications. Extrinsic values can be invoked to motivate indirect moral consideration (see §1) and ascribe partial moral status, such as respect for the patient's biographical life, for the patient's relatives, medical duties or other non-intrinsic factors (Gazzaniga, 2005). However, if we take seriously the ontological-ethical separation between human being and person, with the cessation of the person as the substantial sortal that identifies and determines our persistence conditions, there is arguably no person truly entering a vegetative state. On a theoretical and ethical level (not juridical),

9 While some might consider these examples as exaggerated, recent developments highlight the practical relevance of such ethical considerations. A recent New York Times opinion piece by Jauhar, Patel, and Smith proposed expanding the definition of death to include "irreversibly comatose" patients, specifically to address organ donation shortages (Jauhar et al., 2025). This proposal demonstrates that these are not mere thought experiments but real-world issues arising from personist approaches combined with utilitarian perspectives on death, as critically observed by J.J. Fins in his response to the article (Fins, 2025). Such debates underscore the complex interplay between philosophical concepts of personhood, medical ethics, and practical healthcare challenges.

the previously existing person could lose any moral authority over the biological body and the content of their prior wishes (to sustain life, withdraw treatment, euthanasia, etc.).

The animalist position is quite different from that of personism regarding the identification of entities through a sortal approach, since the substance sortal is the animal. However, the theoretical framework of the person is the same. On an ethical level, if a defender of animalism grants full moral status to the person, once the person ceases to exist as a phase sortal, there are two possible lines of consequences that can follow: (a) the human animal, as soon as it ceases to be a person, loses grounds for moral consideration, thus opening up possibilities for the use (and potential abuse) of the biological body, as previously outlined (in the absence of other extrinsic motivations to protect it); (b) the human animal, despite the irreversible loss of the sortal "person", remains the object of indirect moral consideration for other reasons with partial moral status (as in some versions of personism, see above).

The disjunctivist hybrid approach and the constitution view propose a less rigid view of personism and animalism, thus allowing the theory to be ontologically more "adaptable" in certain complex and ambiguous situations, such as the vegetative state. However, given that both theories rely on a functional conception of person, similar moral potential consequences may apply. As for the hybrid view, since the reference to the concept of person remains functionalist, one could assume the need for direct or greater moral consideration when the psychological phase sortal is present. In such a case, the VS patient does not possess the phase sortal "person", so the possibilities and ethical consequences highlighted by animalism persist. As for the constitution view, once the substantial sortal "person" is lost, there is no reason to ascribe to the living human being any special moral status, since the sameness of the living human organism is not sufficient for the sameness of the person. In other words, the constitution view maintains the possibilities and ethical consequences highlighted with personism.

Ontological personalism is the only theory presented here that is based instead on the neo-Thomist substantialist concept of person. This approach seems to reconcile both the intuitions that a VS patient persists as a human person, while simultaneously being different from their prior self (however, without a change in fundamental nature). In this sense, the theory argues that the statement "is no longer a person" refers to the change of qualities and mental properties that certainly characterize the individual (mainly personality and selfhood, but distinct from personhood itself), but these do not determine the criteria for existence, persistence, and reidentification. From an ethical perspective, the VS patient continues to persist as a person, that is, as an individual corporeal entity of rational nature, so there is no loss of direct moral consideration or full moral status in general.

However, proponents of the functionalist conception of personhood might argue that the persistence of the person entails at least two problematic ethical consequences: the first pertains to issues of justice and fair allocation of resources. If VS patients possess full moral status as persons, it may become problematic to differentiate and prioritize moral obligations between them and patients with more acute conditions involving evident suffering (e.g., acute pain). A strict ontological personalism might have tough time with such morally relevant distinctions. The second is that ontological personalism might entail the risk of slipping into therapeutic adventurism, i.e., an excessive emphasis on cure over care, and into a total non-disposability of life, necessitating the prolongation of the patient's existence at all costs, regardless of poor prognosis or quality of life considerations.

Nonetheless, these two objections would stem from a perhaps superficial interpretation of the ontological personalist view. Regarding the first, it derives from a confusion between possessing full moral status and the resulting moral obligations. Granting a VS patient full moral status as a person does not necessarily entail obligations of beneficence so stringent as

to become excessively burdensome (acting in that way would be supererogatory).[10] Moreover, considering all patients equally as persons does not equate to placing all clinical conditions on the same level. Where a patient is consciously experiencing acute pain, the resources employed and therapeutic treatment will differ from those for a chronic vegetative case, based on appropriate medical criteria. What ontological personalism demands is a "decent minimum" level of care and assistance as a necessary (though not sufficient) condition for all persons. The second objection conflates the substantialist notion of personhood with the idea of the sanctity of biological life and confuses ontological personalism with a vitalistic approach. In fact, as Eberl (2020) argues, the continuation of a human person's existence as a bearer of full moral status is a *prima facie* good. At the same time, avoiding unnecessary suffering and disproportionately burdensome interventions is equally required in order to respect the dignity of the person. In such cases, palliative care procedures are especially appropriate, and it may be morally justified to withhold or withdraw therapeutic interventions that are deemed disproportionate (Eberl, 2003).

5. Pragmatic Considerations from a Personalist Approach to Disorders of Consciousness

Certain conceptions of the person can influence moral agency, especially when certain metaphysical claims become part of common sense. With respect to theories that tend to underestimate or eliminate the moral status of VS patients, duties of respect for the individual regardless of the state of the patient can prevent or lower attitudes contrary to good clinical practice and acts of non-functional de-humanization (e.g., deindividuation, dissimilarity [Haque & Waytz, 2012]); on the other hand, with respect to theories that tend to consider the VS patient as a person of full moral status, cost-benefit analysis can provide useful information for clinical decisions in situations of scarce medical resources.

In the preceding section, I have illustrated the various metaphysical and ethical issues that the functionalist approach faces regarding the vegetative state condition. In contrast, although also ontological personalism may raise potential critiques, it appears to be the most plausible and comprehensive theory in addressing the theoretical challenges posed by this liminal case. Now, moving beyond pure theory and recognizing the situation of epistemic and ethical uncertainty such that of VS diagnoses, I believe that an ontology of the person that favors the attribution of moral status to these patients is preferable also on a pragmatic level (beyond its metaphysical consistency). If variables such as geographic and cultural background, as well as religious beliefs can influence moral considerations and clinical decisions regarding patients (Demertzi et al., 2011), then it is reasonable to think that one's internalized conception of personhood could similarly shape attitudes and approaches in complex clinical situations like the vegetative state.

Ontological personalism may result useful in improving certain attitudes and approaches to complex clinical and existential situations such as the vegetative state (Clark et al., 2025). Considering this, a clinical education based on a specific ethical perspective on personhood could influence – if not properly orient – some attitudes and habits of healthcare professionals towards patients. A personalist approach may encourage recognizing the patient's intrinsic moral status, regardless of external obligations, relations, or extrinsic values. This orientation extends beyond single actions, promoting the development of

10 A supererogatory act is a morally good, voluntary action that goes beyond duty, is not required, and typically benefits others, making its performance praiseworthy while its omission would not be considered morally blameworthy. For a discussion of the concept of supererogatory acts, see Grigoletto (2019). Consider a case of a caregiver or a nurse of a patient in a persistent vegetative state who might spend extra hours each day – or even nights – providing physical care, sensory stimulation, and attentive presence beyond what is professionally required. This would be an instance of a supererogatory act in the medical-care context.

a proactive and lasting ethical-professional habits towards patients with disorders of consciousness. Such a personalist ethics can increase attentiveness to new rehabilitative therapies and avoid numerous epistemic and ethical risks, such as therapeutic nihilism, prognostic pessimism, and self-fulfilling prophecy biases, by acknowledging these patients as vulnerable human persons rather than mere biological entities that need extrinsic values to gain some indirect moral consideration. However, it is important to avoid conflating respect for inherent personal dignity, as advocated by the personalist approach, with excessively empathizing with the patient. The latter, indeed, risks caregiver and healthcare professional burnout from an over-involvement with vulnerable patients' difficult situations, compounded by the multidimensional burdens they face when caring for patients with disorders of consciousness (Leonardi et al., 2023).[11]

REFERENCES

Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. New York, NY: Cambridge University Press;

Blumenthal-Barby, J. (2024). The End of Personhood. *The American Journal of Bioethics*, 24(1), 3-12. doi: 10.1080/15265161.2022.2160515;

Clark, T., Edgley, A., & Kerry, R. (2025). Making Healthcare Decisions on Behalf of People in a Disorder of Consciousness. A "Risk-Making" Theory of Decisional Practices. *AJOB Neuroscience*, 16(3), 129-145. doi:10.1080/21507740.2025.2464112;

Darwall, S. L. (1977). Two kinds of respect. *Ethics*, 88(1), 36-49. doi:10.1086/292054;

DeGrazia, D. (2005). *Human Identity and Bioethics*. New York, NY: Cambridge University Press;

DeGrazia, D. (2008). Moral Status As a Matter of Degree? *The Southern Journal of Philosophy*, *46*(2), 181-198. doi:10.1111/j.2041-6962.2008.tb00075.x;

Demertzi, A., Ledoux, D., Bruno, M. A., Vanhaudenhuyse, A., Gosseries, O., Soddu, A., Schnakers, C., Moonen, G., & Laureys, S. (2011). Attitudes towards end-of-life issues in disorders of consciousness: A European survey. *Journal of Neurology*, *258*(6), 1058-1065. doi:10.1007/s00415-010-5882-z;

Eberl, J. T. (2003). Aquinas on euthanasia, suffering, and palliative care. *The National Catholic Bioethics Quarterly*, *3*(2), 331-354. Doi:10.5840/NCBQ20033257;

Eberl, J. T. (2020). *The nature of human persons: metaphysics and bioethics*. Notre Dame, IND: Notre Dame Press;

Fins, J. J. (2025, August 17). The Debate Over Defining Death. *The New York Times*. Retrieved from https://www.nytimes.com/2025/08/17/opinion/defining-death-organ-donors.html;

Gazzaniga, M. S. (2005). *The ethical brain: The Science of Our Moral Dilemmas*. Washington, DC: Dana Press;

Giacino, J. T., Katz, D. I., Schiff, N. D., Whyte, J., Ashman, E. J., Ashwal, S., Barbano, R., Hammond, F. M., Laureys, S., Ling, G. S. F., Nakase-Richardson, R., Seel, R. T., Yablon, S., Getchius, T. S. D., Gronseth, G. S., & Armstrong, M. J. (2018). Practice guideline update recommendations summary: Disorders of consciousness. *Neurology*, *91*(10), 450-460. doi: 10.1212/WNL.0000000000005926;

Grigoletto, S. (2019). *Only Through Complexity. Morality and the Case of Supererogation*. Padua: Padua University Press;

Haque, O. S., & Waytz, A. (2012). Dehumanization in Medicine: Causes, Solutions, and Functions. *Perspectives on Psychological Science*, *7*(2), 176-186. doi:10.1177/1745691611429706;

11 I would like to thank Jason T. Eberl and Simone Grigoletto for their helpful comments on an earlier draft of this paper.

Jauhar, S., Patel, S., & Smith, D. (2025, July 30). Donor Organs Are Too Rare. We Need a New Definition of Death. *The New York Times*. Retrieved from https://www.nytimes.com/2025/07/30/opinion/organ-donors-death-definition.html;

Jaworska, A., & Tannenbaum, J. (2019). Persons and Moral Status. In A. LoLordo (Ed.), *Persons: A History* (pp. 334-362). New York, NY: Oxford University Press. doi:10.1093/OSO/9780190634384.003.0014;

Kavanaugh, J. F. (2001). *Who Count as Persons? Human Identity and the Ethics of Killing*. Washington, DC: Georgetown University Press;

Kondziella, D., Bender, A., Diserens, K., van Erp, W., Estraneo, A., Formisano, R., Laureys, S., Naccache, L., Ozturk, S., Rohaut, B., Sitt, J. D., Stender, J., Tiainen, M., Rossetti, A. O., Gosseries, O., Chatelle, C., & the EAN Panel on Coma, Disorders of Consciousness (2020). European Academy of Neurology guideline on the diagnosis of coma and other disorders of consciousness. *European Journal of Neurology*, *27*(5), 741-756. doi:10.1111/ene.14151;

Langford, S. (2014). On What We are and How We Persist. *Pacific Philosophical Quarterly*, *95*(3), 356-371. doi:10.1111/PAPQ.12035;

Laureys, S., Celesia, G. G., Cohadon, F., Lavrijsen, J., León-Carrión, J., Sannita, W. G., Sazbon, L., Schmutzhard, E., von Wild, K. R., Zeman, A., & Dolce, G. (2010). Unresponsive wakefulness syndrome: A new name for the vegetative state or apallic syndrome. *BMC Medicine*, *8*(68). doi:10.1186/1741-7015-8-68;

Leonardi, M., Sattin, D., Cacciatore, M., Ippoliti, C., Barbadoro, F., & Magnani, F. G. (2023). Taking Care of Patients with Disorders of Consciousness: Caregivers' Burden and Quality of Life. In: C. Schnakers, S. Laureys (Eds.), *Coma and Disorders of Consciousness* (pp. 221-241). Cham: Springer. doi:10.1007/978-3-031-50563-8_10;

Levy, N., & Savulescu, J. (2009). Moral significance of phenomenal consciousness. *Progress in Brain Research*, *177*, 361-370. doi:10.1016/S0079-6123(09)17725-7;

Mcmahan, J. (2010). Death, Brain Death, and Persistent Vegetative State. In H. Kuhse & P. Singer (Eds.), *A Companion to Bioethics: Second Edition* (pp. 286-298). Hoboken, NJ: Wiley-Blackwell. doi:10.1002/9781444307818.ch25;

Noonan, H. W. (2021). Personal Identity and the Hybrid View: A Middle Way. *Metaphysica*, *22*(2), 263-283. doi:10.1515/MP-2020-0007/HTML;

Olson, E. T. (1997). *The Human Animal: Personal Identity Without Psychology* (Vol. 60, Issue 2). New York, NY: Oxford University Press;

Olson, E. T. (2007). *What Are We? A Study in Personal Ontology*, New York, NY: Oxford University Press. doi:10.1093/ACPROF:OSO/9780195176421.001.0001;

Schnakers, C., & O'Brien, K. (2023). Behavioral Assessment and Diagnosis of Disorders of Consciousness. In: C. Schnakers, S. Laureys (Eds.) *Coma and Disorders of Consciousness* (pp. 17-33). Cham: Springer. doi:10.1007/978-3-031-50563-8_2;

Singer, P. (2011). *Practical Ethics*. New York, NY: Cambridge University Press;

Teichman, J. (1996). Human Beings and Persons. In *Social Ethics: A Student's Guide* (pp. 29-40). Cambridge, Mass.: Wiley-Blackwell;

Tooley, M. (1972). Abortion and infanticide. *Philosophy and Public Affairs*, *2*(1): 37-65;

Wiggins, D. (2001). *Sameness and Substance Renewed*. New York, NY: Cambridge University Press;

Young, M. J., Bodien, Y. G., Giacino, J. T., Fins, J. J., Truog, R. D., Hochberg, L. R., & Edlow, B. L. (2021). The neuroethics of disorders of consciousness: A brief history of evolving ideas. *Brain*, *144*(11), 3291-3310. doi:10.1093/brain/awab290.

SECTION

03

SECTION 3

GENDER, INTERSECTIONALITY, AND THE RECONSTRUCTION OF IDENTITY

E. DÍAZ-LEÓN
Universitat de Barcelona
ediazleon@ub.edu

GENDER IDENTITY, FIRST-PERSON AUTHORITY, AND BELIEF

abstract

Talia Bettcher (2009) has defended a principle known as "The First-Person Authority" (FPA) of gender avowals. *Gender avowals are a specific kind of speech acts, namely, present-tense gender ascriptions from the first-person point of view (which we can also call* gender self-identifications*). This principle has given rise to an interesting debate about what is the best way of articulating and defending the principle in a manner that is useful for promoting the rights of trans people. In this paper, I argue that an adequate explanation of the epistemology of gender self-identifications involves the insight that our knowledge of our attitudinal mental states is based in part on* introspection *of the conscious manifestations of these very dispositional mental states. That is, it is part of the nature of propositional attitudes such as beliefs and desires that they can be* manifested*, under certain manifesting conditions, in the form of* conscious *mental states. Thus, given the strong connection between a disposition and its manifestations, the fact that we can know some of the dispositions' manifestations by introspection explains why I can have an epistemic advantage over other people who are not able to know the conscious manifestations of my mental attitudes by introspection.*

keywords

gender identity, first-person privilege, self-knowledge, amelioration

Phenomenology and Mind, n. 29 - 2025, pp. 114-123
DOI: 10.17454/pam-2910
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

I. Bettcher's account

Talia Bettcher (2009) has defended a principle known as "The First-Person Authority" (FPA) of gender *avowals*. Gender avowals are a specific kind of speech acts, namely, present-tense gender ascriptions from the first-person point of view (which we can also call *gender self-identifications*). This principle has given rise to an interesting debate about what is the best way of articulating and defending the principle in a manner that is useful for promoting the rights of trans people.

Bettcher (2009) argues that the FPA should be understood as an *ethical* requirement rather than an *epistemic* requirement. Her main argument is that the claim that we have epistemic authority about our gender avowals is strictly speaking false. Her reason for this claim is that we do not have first-person epistemic authority about many of our claims about our own mental lives.

She distinguishes, following other philosophers of mind, between "fleeting" mental states, also known as *occurrent* mental states, and "durable" mental states (also known as *standing* mental states) (2009, p. 99). Occurrent mental states are those that appear in the stream of consciousness at a specific time, have a duration, and typically disappear at a later time (think, for example, of a toothache, an itch, the taste of a hot chocolate, or the smell of a rose). These mental states are always conscious. Standing mental states, on the other hand, do not necessarily enter into the stream of consciousness, and can have a much longer duration. For example, many of our beliefs are standing mental states in this sense: I have had the belief that Paris is in France for many years, and will probably have it for many more, even in those times when I am not consciously thinking about Paris nor France. Standing mental states do not have to be conscious: arguably, they are *dispositional* states that have conscious phenomenal states as some of their manifestations. Desires (and other propositional attitudes such as fears and hopes) are also arguably standing states: for instance, someone can have a long-standing desire to be a University professor. This desire can be manifested in the form of behavior (such as hard work) and other mental states (including conscious states such as feeling happiness and relief when one gets a job offer, etc.).

Bettcher argues, drawing on a consensus in philosophy of mind, that although self-ascriptions of *phenomenal states* seem very close to being *privileged*, we cannot be so certain about our self-ascriptions of *standing mental states*.

Let's examine occurrent mental states first. Indeed, many philosophers of mind reject the Cartesian idea that we have epistemic privilege about our own phenomenal states, that is, they deny that we cannot be mistaken about our own mental states, and that no one can know

about our phenomenal states better than ourselves. This *epistemic privilege* can be understood as the combination of two claims: (1) *infallibility*: if I am in mental state M, I *know* that I am in M; and (2) *incorrigibility*: if I *believe* that I am in M, then it is *true* that I am in M. The second claim is the most important one for our purposes since what is more politically significant is whether first-person gender *judgements* are incorrigible or not, that is, if my belief about my gender self-identification could be wrong, and if someone could correct me about it.

Bettcher argues, in the first place, that with regard to our judgements about our mental states, we are pretty certain about our first-person (present-tense) ascriptions of *phenomenal states*: "avowals of [phenomenal] states generally exhibit an immunity that many judgements do not" (100). Strictly speaking, we *could* be wrong about these judgments: given most accounts of phenomenal states, I could have the belief that I am in pain even if I am not really in that state (because there could be some cognitive mechanism that could be blocking or interfering with the correct functioning of my introspective mechanisms, such as bias, or trauma). For example, I might believe that I am in pain when in fact I am just extremely cold. Or in some contexts, I might mistake the feeling of extreme heat with the feeling of extreme cold. But as Bettcher explains, drawing again on a consensus in philosophy of mind, these cases seem to be highly unusual, so that for practical purposes, we can assume that first-person present-tense avowals of phenomenal states are virtually incorrigible. It would be very weird, from an epistemic point of view, for someone to challenge my judgements about my own phenomenal states. It is true that I am not 100% certain, but certainty is not required for knowledge nor strong justification.

Furthermore, even if we accept that it is possible that my avowal of my own phenomenal state is false, we can still accept the following. Imagine that after my *asserting* that I have a phenomenal state, another person *asserts* that I am wrong, or doubts that my judgement is right. This seems to generate a strong epistemic duty from the latter speaker, namely, the speaker should provide some reasons for why the former subject's own phenomenal judgement is wrong. That is, the general fact that we are not 100% certain about our phenomenal judgements does not seem enough justification for someone else to *make the assertion* that my own judgement that I am in a given phenomenal state is not true. And it is difficult to think about what reasons could possibly be available to a third person but not available to the subject of the phenomenal state itself, at least in normal conditions.

Hence, I agree with Bettcher that we can treat phenomenal avowals as virtually incorrigible, at least for practical purposes, in the sense that the default option should be to trust phenomenal avowals, in normal conditions. We could add that in some exceptional circumstances, the hearer of a phenomenal avowal who then goes on to assert that the avowal is not true could possibly discharge their epistemic duties. That is, the speech act of doubting my phenomenal avowal would be an *appropriate* speech act only if the speaker is able to present very strong reasons for why the first-person phenomenal judgement might be incorrect, and *in addition*, why the hearer is in a better position to know than the subject herself. Again, as Bettcher argues, this is not inconceivable (for example, the subject might be under the effects of drugs that could be blocking their introspective mechanisms, and all of this might be known by the hearer), but this is unusual.[1] For this reason, it seems that to deny

1 It could be argued that our introspective judgements about our own phenomenal states are not as reliable as I am assuming (see Schwitzgebel, 2006). In any case, my main claim regarding phenomenal avowals is that it is often inappropriate for a third person to doubt a phenomenal avowal, even if these are not *incorrigible* strictly speaking. In addition, given that Bettcher is happy to accept the incorrigibility of phenomenal avowals, as opposed to the incorrigibility of avowals of standing mental states, I am happy to assume the former for the sake of the argument and focus on the discussion of the latter.

a phenomenal avowal is in most contexts an *inappropriate* move in a conversation, given the very nature of phenomenal states and our knowledge about it (namely, by means of directly introspecting our own conscious phenomenal states).

Let's now examine the alleged incorrigibility of our avowals of *standing* mental states. As Bettcher argues, there are no good reasons to claim that our judgements about our own standing mental states are incorrigible, not even with respect to present-tense avowals: "In attitudinal cases, FPA is less strong" (p. 100). And the main reason is that our knowledge of standing mental states is not by introspection alone. How we come to know our standing mental states is a complex subject, but a plausible position that is widely endorsed has it that knowledge of those mental states is achieved by means of inferential mechanisms, drawing on several pieces of evidence, including introspection of possible conscious manifestations of those dispositional mental states, observations of manifested behavior, past patterns of behavior, other standing mental states about which the subject could have independent reasons for having, and so on.[2] Hence, our knowledge of our standing mental states is based on inferential reasoning, not on introspection alone. Clearly, inferential reasoning on the basis of observable behavior is less reliable than introspection. Still, we can assume for our purposes here that we can have strong justification of this sort for our judgements about our own standing mental states, such as beliefs and desires. However, this falls short of *incorrigibility*, even the slightly modified sort of incorrigibility that we argued above is plausible to ascribe to phenomenal judgements (that is, the default trustworthiness of phenomenal avowals, given the primacy of introspection with respect to third-person epistemic methods concerning phenomenal experience), since this is not so plausible with regard to standing mental states. So, if our judgments about our standing mental states are not incorrigible in any interesting sense, the crucial question becomes whether we still have any interesting kind of *epistemic advantage* from the first-person point of view, compared to the third-person point of view.

Bettcher argues that there is no interesting sort of *epistemic* advantage from the first-person point of view. And the main reason (as I reconstruct her argument) is that given the way we come to know about our own *standing* mental states, the first person and the third person points of view seem pretty analogous, with regards to the forms of access.

Here, we should be careful. It has been established that our judgements about our own standing mental states are not *incorrigible* in any substantive sense, since we could be mistaken about them to the same extent as we are mistaken about many other judgements we make about the external world. But this gives rise to the crucial question of whether we have any epistemic *advantage* from the first-person point of view in order to know my own beliefs and desires, as opposed to the third person point of view, even if it does not amount to incorrigibility. It is clear that first-person judgements about my own standing mental states can be wrong, but can we say that first-person judgements have an *epistemic advantage* with respect to third-person judgments when it comes to knowing a person's attitudinal mental states?

As we said above, the inferential reasoning by means of which we come to know about our (standing) beliefs and desires is based on a combination of different pieces of evidence, including introspection of any conscious manifestations of these dispositional states that there might be. And when it comes to the introspection of our own phenomenal states, it seems clear that we do have an epistemic advantage from the first-person point of view as opposed to the third person, due to the mechanisms of introspection. As I put it above, phenomenal avowals are virtually incorrigible for this reason, since it would be inappropriate in most

2 See Gertler, 2011 for a useful overview.

contexts, for other people, to correct my avowal of a phenomenal state, unless they discharge their epistemic duty to offer sufficient reasons for making such an unusual assertion. The crucial question, then, is whether we can say something similar about avowals concerning standing mental states such as beliefs and desires.

Bettcher claims that we cannot talk about *epistemic advantage* in any substantive sense regarding standing mental states. In response, I think that we are justified in positing at least some sort of epistemic advantage from the first person point of view as compared to the third person, given the fact that the inferential reasoning that we use to come to know about standing mental states also relies in part on *introspection* (of conscious manifestations of our attitudinal mental states), and we do have some first-person epistemic advantage regarding introspection. Of course, Bettcher is right that this does not entail incorrigibility in any interesting sense: there is no sense in which we cannot be mistaken about our standing mental states, from the first-person point of view, as opposed to the third person point of view.[3]

However, I wish to argue that there is a sense in which it is still *inappropriate* for an external person to challenge or doubt my avowal of a standing mental state, unless they supply sufficient additional evidence that overrides my epistemically privileged access to some partial evidence for my avowal, namely, my introspective judgements. It could be the case that someone has evidence based on my observable behavior (about which another person could have more reliable records than myself, for instance), so that this overrides the partial introspective evidence that I have about whether I have a given belief or desire. Still, it can be argued that in order to challenge a first-person avowal of a belief or a desire, an external person has the *epistemic* duty to present sufficient overriding evidence. And this additional duty is precisely what explains the degree of epistemic advantage that the first-person point of view does have concerning our own standing mental states, in my view. That is, if someone were to doubt or challenge a *third-person* claim about someone else's attitudinal states, the former would have the standard epistemic duty to be able to present reasons for their assertion, but in the case of doubting or challenging a *first-person* claim about the subject's own attitudinal mental states, the speaker who challenges this claim would incur in an *additional* epistemic duty to offer reasons for why the doubter knows better than the subject herself, in this particular case.

However, this might not be enough to talk about first-person epistemic *authority* or *privilege* in any substantive sense. Bettcher argues that from an epistemic point of view, avowals of beliefs and desires do not display first-person *epistemic authority*, because they rely on inferential knowledge on the basis of observable behavior, and this is fallible. But still, she argues that it is *wrong* in some substantive sense to challenge or to doubt those avowals. Why? According to Bettcher, the reasons are *ethical* rather than epistemic. And it does not have to do so much with the kind of *evidence* the first person has for their own mental states (which is similar to the kind of evidence available to the third-person), but rather with the kind of *speech act* that we make when we self-report our own standing mental states such as beliefs and desires and other propositional attitudes. Bettcher's crucial claim is that a self-report of a standing mental state is not merely a report of a *fact* (p. 102), for which both the first person and the third person could in principle have similar sorts of evidence, but rather constitutes a different kind of illocutionary speech act, other than reporting a fact.

3 Logue (2021) also argues that our beliefs about our gender identities are not incorrigible, since they can be mistaken. She considers the example of a trans woman who did not believe she was a woman at a certain time before transitioning. Logue argues that it is important to be able to say that she was never a man, and therefore, her belief before transitioning that she was a man was mistaken (2021, p. 130).

This speech act is different in several inter-connected ways. First, when one self-reports a mental attitude such as a belief or a desire, one is *taking responsibility* for those attitudes. Bettcher argues that propositional attitudes can be appraised or criticized, for being true or false, rational or irrational, appropriate or inappropriate. And when a subject reports her own mental attitudes, she is not just describing a certain fact of the matter, given her evidence, but she is also putting her own attitudes "out there" to be publicly scrutinized, so that she can be subject to praise or criticism because of their rationality or appropriateness. Because of this, there is an important asymmetry between first-person and third-person attitude reports. And Bettcher argues that it is wrong for someone to assume that they are allowed to put someone else's attitudes out there for public scrutiny. This is something that (at least in normal circumstances) only the owner of the attitudes is entitled to do. As Bettcher explains, our mental attitudes are *private* in the sense that it is wrong for others to make them public without our consent. For this reason, to claim that one has expertise about another person's mental attitudes is ethically problematic, precisely because we are making something public that is "none of our business", that is, only the first person is entitled to make them public. Hence, this explains the special illocutionary force of attitudinal self-reports: we are making our attitudes public, attitudes that other people should not make public without our consent, and furthermore, we are putting our own attitudes out there for public scrutiny, to be appraised or criticized, in a way that makes us *responsible* for their merits or flaws. This is the sense in which the first person is an *authority* about their own mental attitudes: not because she has *epistemic* expertise over them, but because she is *authorized* (unlike external subjects) to make them public and to take responsibility for them in making them subject to rational scrutiny by others. In other words: we are entitled to becoming responsible for our own attitudes, but other people are not entitled to make us responsible for our own attitudes.[4]

Bettcher (2009) advocates for a change of the meaning of the terms 'man' and 'woman' (in particular, for a shift in our patterns of usage of such terms and cultural practices involving these terms). According to Bettcher, the way of using these gender terms in mainstream linguistic communities is closely related to the mainstream cultural practices about gender presentation. In the same way that gender presentation (e.g. clothes, hairstyle, etc.) is supposed to represent genital status, the ascription of gender terms such as 'man' and 'woman' are also supposed to represent people's genital status. But Bettcher argues that genital status should remain *private*, that is, it is wrong for others to expect us to reveal something that is not really "their business" to know. For these reasons, we should change both norms of appearance surrounding gender presentation and norms of use regarding gender terms. We should change the meaning (and use) of gender terms so that they do not aim to reveal genital status but rather are used to communicate someone's gender *self-identification*. This clearly respects the First-Person Authority principle about gender, according to Bettcher. Let me explain why: If gender terms aim to refer to someone's self-identification, then when we talk about someone's gender, we talk about how they self-identify. And it is clear that claims about how one self-identifies display first-person authority in the ethical sense: it would be wrong for someone else to claim expertise about our own gender self-identification. This is not due to the fact that the subject could never be wrong

4 This claim by Bettcher has intuitive appeal, although on reflection, there seem to be some problems lurking. For example, we might be entitled to make other people responsible for their propositional attitudes, such as sexist or racist or homophobic propositional attitudes. Furthermore, there might be exceptions to the restrictions to make my attitudes public: for example, my therapist might be *authorized* to make some of my hidden attitudes public (to me). Or a very close friend of mine might be authorized to tell me about some of my hidden attitudes, such as my homosexual sexual desires which I want to repress due to internalized homophobia. Thus, there are exceptions where other people might also be allowed to make someone else's attitudes public and subject to scrutiny.

about her own self-identification (because she could be wrong), but rather because her *speech act* of declaring her gender is thereby a speech act of declaring her gender self-identification (assuming gender terms come to refer to gender self-identification), that is, a gender avowal on this understanding becomes an act of the subject making public, and taking responsibility for, the mental attitude of self-identifying with a gender. And as Bettcher argues, other people are not *allowed* to disclose my own mental attitudes without my consent. This is what the ethical reading of first-person authority involves.

Now, this argument still gives rise to an epistemic worry, as follows. Bettcher has clearly established that it is ethically problematic for another person to *claim* that they have expertise about my gender self-identification, because this denies my autonomy, in the sense that it violates an ethical principle according to which other people are not allowed to reveal something that is not up to them to reveal. Nonetheless, could it be argued that someone else could have a *belief* about my own gender self-identification that is epistemically better justified than my own belief about it? Bettcher's argument concerns *assertions* about someone's gender, and she argues that the *speech act* of gender avowal is different from the speech act of revealing somebody else's gender, so that only the former speaker, but not the latter, is *authorized* to perform such a speech act.[5] This ethical argument concerns *speech acts*.

But what about the justification of our *beliefs*, in cases these beliefs are not linguistically expressed? The central question that arises is whether my first-person *beliefs* about my attitudinal mental states have any *epistemic advantage* over third-person beliefs about someone else's mental states. It is not clear to me what Bettcher's argument would entail about this issue. Since this is an important question, more attention should be paid to this issue.[6] In the next section I will explore an alternative account that offers an answer to this question, an answer that I will argue is misguided. And in section III I will develop what I take to be a more satisfactory answer.

II. Fileva's account

Iskra Fileva (2021) has recently offered a novel account of gender, according to which the features that constitute gender are all knowable by introspection. If this account were correct, we would have an answer to the question I posed in the previous section, to wit: if gender can be known by introspection, then this explains why my *belief* about my own gender is *epistemically privileged* as compared to other people's beliefs about my own gender. As we saw, on Bettcher's own account, gender is based on attitudes of self-identification. And since, according to Bettcher, my beliefs about my own attitudinal mental states are not epistemically privileged, as compared to other people's beliefs about my own attitudes (for we all come to know about my own attitudes in the same way, that is, by inferential reasoning), then it follows that my beliefs about my gender are not epistemically privileged. This is why Bettcher defends the FPA principle as an *ethical* claim, not as an *epistemic* principle. But Fileva disagrees: she claims that our gender is *constituted* by aspects of our experiences that are knowable by

5 Does this mean that no-one else is authorized to use gender pronouns and gender terms to reveal my gender identity? This could be a possible consequence of Bettcher's ethical reading of the FPA, but this is not how I am understanding it. Bettcher's view, at its core, is that a first-person gender avowal is a speech act of making the subject's own attitudes public, whereas a third-person gender ascription is merely the speech act of describing a fact, which can be true or false, justified or unjustified, and so on. This is the asymetry between the first person and the third person: it is not epistemic, but rather ethical, having to do with the evaluation of what the speaker is doing, from a normative point of view (i.e. describing someone's attitudes vs. making my own attitudes public).

6 Turyn (2023) argues that we have ethical reasons to advocate for an epistemic version of FPA. One main reason is that since it is possible to wrong someone because of what we believe about them, it can be argued that to fail to believe someone's gender avowal can constitute a wrong, and therefore, we ought to believe what others say about their gender identities.

introspection. Her account is a two-tier account. First, she poses a necessary and sufficient condition for someone to have a gender: S has gender G iff S would identify with gender G in ideal circumstances (for example, free from oppression) (2021, pp. 189-190). But she acknowledges that circumstances are not always ideal. Then, she wonders what features of our experiences would give rise to our identifications with a certain gender in ideal circumstances, and she claims that the facts of experience that give rise to gender identification include the following: (a) "comfort or discomfort with the sexual characteristics of one's body" (p. 190); (b) identification with and affinity for people of the sex you identify with (for example, ciswomen would identify with people of the same sex, transwomen would identify with people of a different sex than the sex they were assigned at birth, etc.); (c) " a substantial number of ... preferences about clothing and appearance" (p. 190); and (d) "a strong and persisting preference for being recognized and treated as a man or a woman" (p. 190).

Fileva realizes that none of these criteria is *individually necessary* since there can be people who identify as, say, women, but do not satisfy some of these criteria. For example, many women do not have a preference for clothing and appearance associated with being a woman but can actually have a strong preference for clothing and appearance associated with being a man (as in the case of butch lesbians, or gender non-conforming women). Also, many women do not have a strong affinity with other women but rather prefer the company of men. Fileva responds to this worry by claiming that criteria (a-d) should be understood as a *family-resemblance concept* rather than in terms of necessary and sufficient conditions. In particular, Fileva claims that none of the criteria (a-d) is individually necessary, and furthermore, she adds, "the question is left open of whether they are jointly sufficient (my account allows for the possibility that they are not, though I consider this possibility highly unlikely)" (2021, p. 194). Also, since the account should be understood in "family-resemblance" terms, one could ask how many criteria are necessary in order to satisfy the account. Fileva answers that this is an empirical question. We can conclude, then, that on this account, satisfying an unspecified number of the criteria (a-d) is necessary, where what counts as a sufficient number of criteria is left open, and where no single criterion is individually necessary. Moreover, satisfying all the criteria may not be jointly sufficient to satisfy the account, although this possibility is deemed highly unlikely.

Unfortunately, I think this account faces some serious worries. My main worry is that there could be women who satisfy none of these criteria (a-d), and they still count as women, so it follows that satisfying a sufficient (even if vague) number of criteria is not necessary. Fileva's central claim is that someone is a woman if and only if, in ideal circumstances, she would identify as a woman. Given that this claim by itself does not make the epistemology of gender any easier (since we are often not in ideal circumstances), she aims to offer additional criteria that would make the epistemology of gender easier, and this is why she offers criteria (a-d). Also, she seeks to address the question of the basis on which we identify with a particular gender such as being a man or a woman. To this, she answers: on the basis of (a sufficient number of) criteria (a-d). As we have seen, Fileva also acknowledges that the criteria (a-d), taken together, are not *jointly sufficient*, that is, someone can satisfy *all* these criteria related to being a woman, say, because of societal pressures, without being a woman (since she would not identify as a woman in ideal circumstances), although this is highly unlikely on her view. Furthermore, she argues that satisfying a sufficient (even if vague) number of these criteria is *necessary* for being a woman, even if it is not sufficient for being a woman.

I disagree: it seems to me that the *disjunction* of criteria (a-d) is not necessary either, that is, someone can be, say, a woman, even if she does not satisfy *any* of these criteria (but would still identify as a woman in ideal circumstances). Given that the criteria are neither jointly sufficient nor disjunctively necessary, it is not clear that this really counts as a family-resemblance account. Rather, criteria (a-d) could be understood as a set of *fallible guides* for

knowing someone's gender: if someone has the experiences related to those criteria, it is likely that they will be of the corresponding gender, but this guide might be fallible in many cases. Fileva agrees that the criteria (a-d) might be fallible, but she also claims that criteria (a-d) provide a family-resemblance account of gender, which as I have argued does not seem correct. Furthermore, she claims that criteria (a-d) "are introspectively accessible aspects of experience that are constitutive of gender and simultaneously underwrite our knowledge of our own gender" (2021: 190). And she adds: "dysphoria or the absence thereof does not simply provide evidence of gender, it is partly *constitutive* of it" (190). In response, it seems controversial that criteria (a-d) are *constitutive* of gender, since those criteria are not jointly sufficient for having a certain gender, and furthermore, someone could have a certain gender without satisfying any of the corresponding criteria, that is, the criteria (a-d) are neither jointly nor disjunctively necessary to have a gender.

For all these reasons, Fileva's account has not adequately explained the epistemology of gender yet. If criteria (a-d) are no more than *fallible* epistemic criteria for ascertaining someone's gender (including my own gender), then what is gender-identification really based on? We can agree that criteria (a-d) are easy to know from the first-person point of view, but criteria (a-d) do not *determine* someone's gender.[7] So, we can still ask: what determines someone's gender, and is *that* knowable by introspection? On Fileva's account, the only criterion that is truly necessary and sufficient for gender is self-identification *in ideal circumstances*. But this is clearly not knowable by introspection either. So Fileva's account does not provide an answer to our question, namely, whether our *beliefs* about our own gender identities are epistemically privileged as opposed to third-person judgements.

III. A possible solution

In my view, Bettcher's account already provides the resources for an adequate explanation of the epistemology of gender self-identifications, including how our first-person *beliefs* about gender identity (in addition to our first-person gender *avowals*) can be considered privileged. The reason was already suggested in sect. I: our knowledge of our attitudinal mental states is based in part on introspection of the *conscious manifestations* of these very dispositional mental states. For it is part of the nature of propositional attitudes such as beliefs and desires that they can be *manifested*, under certain manifesting conditions, in the form of conscious mental states. Beliefs and desires also have other *non-conscious* manifestations such as causing behavior and other attitudinal mental states. But, given the strong connection between a disposition and its manifestations, the fact that we can know some of the dispositions' manifestations by *introspection* explains why I can have an epistemic advantage over other people who are not able to know the conscious manifestations of *my* mental attitudes by introspection.

The central question above was whether we can say that our *beliefs* about our own gender are, in any meaningful sense, *epistemically privileged*—even if we do not express them linguistically, and therefore, they are not subject to Bettcher's ethical FPA principle. My answer to this question is that our standing mental states of identifying with a certain

7 Fileva claims that criteria (a-d) can be known by introspection, but this is not clear. Most of the criteria appeal to *attitudinal* mental states, such as identification and affinity (b), preferences about clothing and appearance (c), and a persisting preference for being recognized in certain ways (d). These all seem to correspond to dispositional mental states rathen than ocurring mental states. Perhaps criterion (a), about comfort or discomfort, may correspond to phenomenal mental states, but it also has a dispositional dimension in my view. Given the view developed in the previous section, these attitudinal mental states can be known by inference on the basis of multiple factors, including (but not only) the *manifestation* of the dispositions in the form of *conscious* mental states, where these conscious mental states are clearly knowable by introspection. In this way, we can be said to have a certain degree of *epistemic priviledge* with respect to criteria (a-d) from the first person point of view, even if these are not completely knowable by introspection alone.

gender can be known only by means of inferential reasoning (which in principle everyone is in a similar position to achieve). Crucially, however, this inferential reasoning is based on a combination of pieces of evidence, including the *introspection* of conscious manifestations of those very attitudinal mental states, as well as the observation of behaviors that are also a manifestation of those mental states.[8]

To conclude, we can agree with Bettcher that we are not incorrigible about our own gender judgements.[9] But, in addition to an *ethical* version of the FPA principle about gender avowals, we can also hold a modest version of an *epistemic* FPA principle, according to which we are not incorrigible, to wit: I could believe that I have gender G, when this is false, but I am typically in a much better position to know my gender identity than other people are with respect to my own gender identity. Thus, a first-person judgement about gender identity is epistemically privileged with respect to a third-person judgement about gender identity. This means that if I were to linguistically express my gender judgement and someone were to challenge it or doubt it, this challenge would come with a strong epistemic duty, that is, not only the usual epistemic duty to be able to provide evidence to justify their judgement (as with most assertions), but also *additional* evidence in order to justify why they have good reasons to believe that their third-person point of view in this particular case *overrides* my first-person point of view. I think this result nicely complements Bettcher's defense of an ethical FPA principle and can be considered as an *epistemic* version of the FPA principle.[10]

REFERENCES

Bettcher, T. M. (2009). Trans identities and first-person authority. In L. J. Shrage (Ed.), *You've changed: Sex reassignment and personal identity* (pp. 98-120), Oxford: Oxford University Press;

Fileva, I. (2021), The gender puzzles. *European Journal of Philosophy*, 29(1), 182-198. DOI: 10.1111/ejop.12550;

Gertler, B. (2011). *Self-Knowledge*. Milton Park: Routledge;

Jenkins, K. (2016). Amelioration and Inclusion: Gender Identity and the Concept of *Woman*. *Ethics,* 126(2), 394-421;

Jenkins, K. (2018). Toward an Account of Gender Identity. *Ergo*, 5(27), 713-744;

Keeling, S. (2021). Knowing our Reasons: Distinctive Self-Knowledge of Why We Hold Our Attitudes and Perform Actions. *Philosophy and Phenomenological Research*, 102, 318-341. DOI: 10.1111/phpr.12655;

Keeling, S. (2023). Accounting for doxastic agency: Mental action and self-awareness. *Synthese* 201, 185. DOI: 10.1007/s11229-023-04069-7;

Logue, H. (2021). Gender Fictionalism. *Ergo* 8(28), 125-162;

Schwitzgebel, E. (2006). The unreliability of naive introspection. *Philosophical Review*, 117 (2), pp. 245-273;

Turyn, G. (2023). Gender and first-person authority. *Synthese*, 201, 122. DOI: 10.1007/s11229-023-04125-2.

8 Moreover, it could be argued that, from the first-person point of view, we have better epistemic access to our own behavior—that is, knowledge of our *actions*—than to the behavior of others, about which we lack a *sense of agency*, a sense arguably accessible through introspection. (See Keeling 2021, 2023.)

9 Jenkins (2018) also agrees: whereas she claimed in her (2016) paper that we do have the genders that we believe we have (i.e. what I called *incorrigibility* above), Jenkins (2018) claims, drawing on Bettcher (2009), that not everyone who believes that they have a certain gender is right about it. Jenkins (2018) endorses the ethical version of FPA, not the epistemic version. In this paper, I aim to clarify to what extent we can defend an epistemic version too.

10 This work has been funded by grant PID2021-124100NB-I00, funded by MCIN/AEI/10.13039/501100011033 (Spanish Ministry of Science and Innovation). I am also grateful to audiences at Glasgow, Lisbon, Sheffield, Wollongong, and the Pacific APA, as well as an anonymous referee and the editors of this volume.

MARTINA GIOVINE
Vita-Salute San Raffaele University
martinagiovine@outlook.it

LANGUAGE AND IDENTITY: THE PSYCHOLOGICAL IMPLICATIONS OF MISGENDERING FOR TRANS* INDIVIDUALS

abstract

This paper examines the psychological and ethical implications of misgendering, understood as a form of microaggression that undermines the gender identity of trans individuals. Drawing on psychological, medical, and philosophical literature, it explores how structural injustices and social gender norms contribute to the vulnerability of trans* people and affect identity formation. While scholars such as Dembroff and Wodak (2018) and Kapusta (2016) have already identified psychological harm as one reason to morally oppose misgendering, this study develops that argument by providing a more detailed analysis of its psychological impact. The paper evaluates Dembroff and Wodak's twofold claim: a moderate duty not to misgender trans* individuals and a radical duty to eliminate gender-specific pronouns altogether. It endorses the moderate claim, while challenging the radical one, arguing that gendered language can play a crucial role in affirming personal identity. Finally, it calls for context-sensitive strategies to prevent misgendering and promote inclusive linguistic practices.*

keywords

misgendering, gender, microaggression, identity, psychological harm

Phenomenology and Mind, n. 29 - 2025, pp. 124-137
DOI: 10.17454/pam-2911
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

Introduction

Social recognition of an individual's gender identity, that is, their internal sense of being male, female, or non-binary,[1] is fundamental to well-being, self-construction, and life planning within a given social context. As Jae M. Sevelius (2013) argues, such recognition is important for affirming a person's sense of self and shaping how they wish to be perceived and treated by others, including family members, friends, and strangers (Sevelius, 2013).

Linguistic practices reflect and facilitate this recognition. For instance, research has shown that using a person's preferred name and pronouns is crucial, serving as distinctive gender identity markers (Gibbs, 2024).[2] Using language that aligns with an individual's gender identity in social interactions is a fundamental act of acknowledgement. This social recognition tends to occur more readily for individuals whose gender identity aligns with the sex assigned at birth and gender expression. However, trans* individuals (both binary and non-binary)[3] are significantly more likely to experience a lack of social acknowledgement of their gender identity (Sevelius, 2013).[4] This is particularly evident when incorrect names and pronouns are used, a practice known as "misgendering".[5]

1 The term "non-binary" refers to a range of gender identities, including those whose gender identity falls between or outside of the traditional male and female categories, those who may identify as a man or a woman at different times, and those who have no experience of having a gender (Matsuno & Budge, 2017).

2 The names and pronouns that people use may not always match those assigned to them at birth, particularly in the case of trans* individuals. To respect a person's identity, people should use their chosen name and the pronouns, even if their identification documents show a different name or gender. Further information can be found in the National LGBTQIA+ Health Education Center's published glossary, available at: https://www.lgbtqiahealtheducation.org/publication/glossary/.

3 Please note that in this work, the term "trans*" generally refers to individuals whose gender identity does not align with the sex they were assigned at birth. For clarification, the sex assigned at birth refers to when a person is categorised as female, male, or intersex based on the appearance of their genitals and/or their sex chromosomes (LGBTQIA+ Health Education Center, 2024). Using the term "trans*" rather than simply "trans" means that it encompasses both binary and non-binary trans* people. To specifically refer to non-binary people, therefore, I will use the term "non-binary trans* people".

4 The term "gender expression" refers to the "external appearance of one's gender identity, usually expressed through behaviour, clothing, body characteristics or voice, and which may or may not conform to socially defined behaviours and characteristics typically associated with being either masculine or feminine" (Human Rights Campaign, s.d.). Further information is available in The Human Rights Campaign's glossary of terms at the following URL: https://www.hrc.org/resources/sexual-orientation-and-gender-identity-terminology-and-definitions?utm_source=chatgpt.com.

5 Deadnaming, i.e., referring to a person by the name they were given at birth rather than the name they

Misgendering is widely recognised as a form of *microaggression*, that is, a practice that conveys hostile, derogatory, or negative insults towards an individual or group of people based on their membership of a social category. Although misgendering may sometimes stem from ignorance rather than overt malice, this does not negate its impact. Microaggressions, whether intentional or unintentional, can cause significant harm (Sue, 2010). The discriminatory nature of such acts has been extensively discussed in psychological and medical literature (see, for example, Sue, 2010; Nadal et al., 2016), and these discussions have also been explored in philosophy. Robin Dembroff and Daniel Wodak (2018), as well as Stephanie Kapusta (2016), offer compelling arguments for the moral contestability of misgendering. They argue that misgendering can cause psychological harm, constitute a moral wrong, and impose political disadvantages, each of which is a distinct reason to oppose this linguistic practice (Kapusta, 2016; Dembroff &Wodak, 2018).

One of the most significant consequences of misgendering is its psychological impact. Studies have shown that it can cause distress, anxiety, and depression in trans* individuals (Kapusta, 2016; Dembroff & Wodak, 2018). As McLemore (2015) observes, these effects are closely linked to how individuals recognise their own identities and how others perceive them. The recognition of a person's gender identity operates at both an interpersonal level and a collective level of social belonging. Therefore, a failure to acknowledge a person's gender identity can have a profound impact on their mental health (for an overview, see McLemore, 2015). This issue highlights the importance of avoiding misgendering trans* individuals.

Although Dembroff and Wodak (2018) and Kapusta (2016) have addressed the ethical dimension of misgendering, including its psychological harm, they do not examine this harm in detail or situate it within the broader framework of microaggressions. This study builds on their arguments by providing a more in-depth analysis of the psychological consequences of misgendering, focusing on its destabilising effects on personal identity formation. To this end, I begin by discussing how structural injustices, such as transphobia and social gender norms, shape identity formation (§1). I then turn to the medical and psychological research that frames misgendering as a microaggression. I pay particular attention to its characteristics, as well as to the distinction between intentional and unintentional acts (§2). Next, I discuss two philosophical claims put forward by Dembroff and Wodak (2018). The first is a moderate claim that we have a negative duty not to commit misgendering towards both non-binary and binary trans* individuals (§3). The second is a radical claim that we have a moral duty to refrain from using gender-specific pronouns for everyone (§4). I conclude by affirming my agreement with the authors' moderate claim and, in contrast, I argue that the radical claim remains highly contentious. In support of this, I present compelling arguments put forward by Hernandez and Crowley (2024), which illustrate the significant role that expressing gender through language can play in affirming personal identity and how this may ultimately challenge and displace the radical claim.

1. Recognition of personal identity and structural injustice

The construction of personal identity is a multifaceted process that unfolds within specific social and cultural contexts. It is closely interwoven with the acknowledgement and the recognition of the gender identities of others. This process is shaped by the social implications inherent in each gender category. Specifically, each gender-based social category implies certain "constraints and enablements" (Jenkins, 2023) and is associated with social roles, expectations, norms and practices, and self-conceptions (Dembroff, 2018,

prefer, is a form of misgendering. Further information can be found in the following document: https://www.lgbtqiahealtheducation.org/publication/glossary/.

p. 24). Moreover, the social positioning of individuals within gender categories is deeply influenced by structures of injustice, such as transphobia. These structures often operate through dominant social meanings attached to gender, producing marginalisation and oppression for individuals perceived as members of certain gendered social classes. For instance, normative assumptions about embodiment, such as the idea that women must have specific biological traits, may hinder trans* women's access to recognition and social resources. Such structural constraints contribute to the continued marginalisation of trans* identities (Jenkins, 2023).

Trans* individuals may encounter various forms of oppression and marginalisation in different areas of society, including terms of societal discourse, legal and economic practices, conceptual limitations, the use of derogatory language, medical barriers, and other social systems. For example, the misrecognition of the existence of non-binary trans* identities due to a binary conception of society is a prevalent conceptual gap. This concept stems from the idea that the feminine and masculine genders represent the full gender spectrum.[6] Structures of injustice, such as transphobia, are manifested through institutional norms, including denying trans* individuals the right to self-identify their gender on official documents, as has been seen in numerous US states.[7] These structures of injustice result in the exclusion and impediment of self-representation (Dembroff, 2018).

Language also influences the potential for self-narration, as well as its limitations and, in extreme cases, the complete negation of personal identity. Unjust social structures, such as transphobia, are perpetuated through linguistic and discursive practices. Misgendering, for example, constitutes an unfair linguistic practice that contributes to the marginalisation of trans* individuals. As Jenkins (2023) argues, the social meanings attached to gender categories can shape the conditions under which people are recognised, and therefore determine whose identities are affirmed and erased. While not focused specifically on language, her account shows how gendered structures of recognition can undermine the self-expression and social legitimacy of trans* individuals (Jenkins, 2023). Furthermore, the trans* community has emphasised the importance of using the correct language for recognising the gender identity of trans* people. They have emphasised that this is not merely a matter of preference or an optional choice: addressing an individual with the wrong pronouns or name can be perceived as disrespectful, harmful, and even dangerous. Failing to respect and acknowledge a person's gender identity can be considered a discriminatory act (Sevelius et al., 2020). Consequently, the use of language can either facilitate or hinder the recognition of gender identity.

Misgendering can therefore be considered a form of oppression that undermines the social existence and self-narration of trans* individuals. Jenkins (2023) emphasises that misgendering is not merely a matter of individual perception, but rather it has concrete effects on one's ability to exist as a recognised social subject and impacts people's agency (Jenkins, 2023; Kapusta, 2016). In the personal sphere, misgendering may therefore conflict with many individuals' identities. Furthermore, in a social context, this linguistic practice may conflict with the social needs of the trans* community.

6 According to Dembroff (2018), non-binary trans* individuals experience systemic misrecognition and neglect. Their marginalisation extends beyond that experienced by binary trans* identities and manifests in various domains, including education and medicine (Dembroff, 2018).

7 The regulations relating to gender declaration vary across different geographical jurisdictions. In the United States, for example, some states prohibit individuals from declaring a gender other than the one assigned to them at birth. In other states, however, the process of changing gender on official documents is subject to rigorous procedural requirements. For example, in some states, certification of surgical interventions is a prerequisite for gender transition, though these interventions are often financially inaccessible to trans* individuals (Dembroff, 2018).

A substantial body of medical and psychological literature has demonstrated that misgendering can invalidate the identity of trans* individuals, causing social and psychological distress and gender dysphoria. "Gender dysphoria" refers to "distress experienced by some people whose gender identity does not align with their sex assigned at birth based on societal expectations; or distress experienced when a person's gender identity and/or gender expression is not affirmed" (LGBTQIA+ Health Education Center, 2024).[8] To examine this phenomenon in greater depth, the subsequent section will review the existing literature on the psychological consequences of misgendering and its connection to the formation of personal identity.

2. Misgendering as a microaggression

Misgendering, i.e., referring to someone using pronouns or names that do not reflect their gender identity, is increasingly recognised as a form of *microaggression* (Rogers, 2021). Derald Wing Sue (2010) defines microaggression as any brief and common verbal, behavioural, or environmental humiliation that communicates hostility towards an individual or group of people based on characteristics such as race, gender, sexual orientation, or religion (Sue, 2010). While originally used to describe forms of racial discrimination, the concept has since been extended to include gender-based forms of exclusion (Arijs et al., 2023).

Microaggressions often go unnoticed by both those who commit them and those who experience them, as they are deeply embedded in normative social practices. They can take many forms, from body language and facial expression to terminology, offhand comments, and discriminatory institutional policies. Trans* individuals, for instance, frequently report experiencing denigrating and dismissive treatment in settings such as healthcare, when acts of misgendering communicate otherness and can cause embarrassment or humiliation. These interactions often stem from misperceptions and dehumanising beliefs that depict gender transition as invalid or unnatural (Nordmarken & Kelly, 2014).[9]

These acts are not isolated but often persistent, shaping the everyday lives of marginalised individuals. As Sue (2010) argues, microaggressions can reflect cultural worldviews saturated with ethnocentric values, prejudices, assumptions, and stereotypes. They sustain binaries such as inclusion/exclusion, superiority/inferiority, desirability/undesirability, and normality/abnormality (Sue, 2010). In the case of misgendering, this process renders stereotypical understandings of gender visible and explicit (Nordmarken, 2014). The use of incorrect pronouns can signal that a trans* person's self-defined identity is not recognised as legitimate, and in the case of non-binary individuals, that their identities are not acknowledged at all.

Crucially, microaggressions may be intentional or unintentional, but both forms produce harm (Sue, 2010). Unintentional misgendering, often resulting from ignorance, lack of exposure to gender diversity, or assumptions based on appearance, is widespread and perceived as less severe. Yet its apparent harmlessness makes it more pervasive and insidious (Sue, 2010). As Kevin L. Nadal (2016) notes, the pervasiveness and the continuous and daily nature of the microaggressive of these experiences contribute significantly to psychological distress. Victims often report gender dysphoria, feelings of validation, and a diminished sense of belonging (Nadal et al., 2016).

There are many reasons why unintentional misgendering occurs. These include

8 The LGBTQIA+ Health Education Center (2024) highlights that the severity of gender dysphoria can vary greatly among trans* and gender diverse individuals (LGBTQIA+ Health Education Center, 2024). Further information can be found at the following URL: https://www.lgbtqiahealtheducation.org/publication/glossary/.

9 It should be noted that Sonny Nordmarken and Reese Kelly (2014) have themselves questioned the term "microaggression" itself, pointing out that the negative impact of individual cases may not always be minimal (Nordmarken & Kelly, 2014).

misunderstanding or misinterpreting the gender identities of the victims, or having insufficient information about their gender (Nordmarken, 2014). As Mylene Gibbs (2024) reports, trans* individuals (whether binary or non-binary) are especially vulnerable when their expression is perceived to match the gender assigned at birth. This misalignment can lead to others incorrectly perceiving a person's gender identity (Gibbs, 2024). Moreover, Kai Jacobsen et al. (2023) highlight that in the case of misgendering non-binary trans* individuals, there is the additional issue that non-binary genders have not yet achieved widespread social acceptance (Jacobsen et al., 2023). Moreover, an empirical study by Quinn Arijs et al. (2023) revealed that non-binary participants experienced a higher prevalence of using incorrect pronouns and deadnaming when their gender expression aligned with expectations for their sex assigned at birth. Furthermore, participants reported that these practices caused dysphoric feelings and resulted in the perception that their identities were being denied (Arijs et al. 2023). Offenders often downplay the act, dismiss the victim's experience, or excuse themselves by citing habit or difficulty in adapting, responses that worsen the psychological toll and obstruct empathetic understanding (Nadal et al., 2016).

Intentional misgendering, though less common, is often more traumatic. Motivations may include transphobia, a refusal to acknowledge trans* identities, or a desire to ridicule and delegitimise. Some testimonies describe perpetrators who deliberately misgender to provoke or incite others to do the same. Others minimise the victim's experience and the harm caused by claiming that using correct names and pronouns requires too much effort. It should be noted that intentional misgendering is often perceived as more severe and violent than unintentional misgendering (Arijs et al., 2023). In both cases, misgendering functions as a mechanism of exclusion and symbolic violence.

Finally, misgendering occurs across contexts, from interpersonal and familiar to institutional and systemic. A common example of interpersonal misgendering is when family members refuse to use the correct pronouns or names of a trans* relative. On a systemic level, misgendering is embedded in institutions that fail to recognise self-defined identities. For instance, systemic misgendering is when bureaucratic systems do not allow individuals to modify their gender markers, such as in documents, or when public spaces are structured according to the gender binary, such as bathrooms reserved exclusively for women and men (Nadal et al., 2016). These systemic practices do not merely reflect social norms but actively reinforce them, denying recognition and perpetuating inequality.

3. The negative duty not to commit misgendering

Given their moral and political implications, the importance of using pronouns that align with an individual's gender identity has been explored in philosophy. Dembroff and Wodak (2018) discuss two philosophical claims: a moderate one and a radical one. The moderate claim asserts that we have a negative duty not to use binary pronouns when referring to non-binary individuals. Notably, this argument is an intermediate step in the authors' broader argument in support of their radical claim, which is addressed in the next section. The moderate claim is supported by an argument from analogy: just as it is wrong to use male pronouns to refer to a trans* woman, so it is likewise wrong to use binary pronouns to refer to a non-binary individual (Dembroff & Wodak, 2018).

In support of this argument, the authors cite the example of Group Captain Catherine McGregor, a trans* woman who was publicly misgendered with the pronoun "he" by Mark Latham, the former leader of the Australian Labor Party. According to Dembroff and Wodak (2018), this act constituted a moral wrong. While there may not have been a positive moral duty to refer to McGregor with the pronoun "she", since alternatives like "Catherine McGregor" or "Group Captain" were available and morally correct, there was nonetheless a negative duty not to refer to her as "he". This reasoning extends by analogy to non-binary

individuals: we are under a negative duty not to refer to them using male or female pronouns (Dembroff & Wodak, 2018).

According to Dembroff and Wodak (2018), the negative duty to avoid misgendering (whether the individual is binary or non-binary) is grounded in the harmful consequences it entails. These consequences include: (i) expressions of disrespect, (ii) denial of resources, (iii) undermining of intelligibility, and (iv) reinforcement of oppressive ideologies. Firstly, (i) misgendering a binary or non-binary trans* person expresses disrespect towards their social identity and denies their gender identity. Misgendering McGregor, for example, is equivalent to disrespecting her gender identity. In other words, misgendering a binary trans* individual communicates disrespect for their identity, and the same goes for a non-binary trans* individual. Furthermore, the authors argue that misgendering can cause severe physical and psychological health problems, whether intentional or not. Secondly, (ii) misgendering implies the potential denial of resources to which one is entitled. For instance, classifying a trans* woman as male means that she cannot apply for women-only positions or access female spaces. Conversely, using binary pronouns for a non-binary person implies that they must fit into a binary organisation that allocates resources based on gender. A third consequence is (iii) the undermining of the intelligibility of actions and choices. Generally, a variety of social norms can be applied to actions and choices with respect to a given social category, such as gender. For instance, the idea that women shave their legs could be considered a gender-specific norm, and individuals can choose whether to adhere to or violate that norm. However, if a trans* woman is referred to with masculine pronouns, these norms are not recognised as applying to her, consequently rendering her unable to violate or adhere to them. This results in the suppression of an important form of expression. The same is true for non-binary trans* people: misgendering can undermine their ability to reject the norms associated with being male or female. Finally, (iv) misgendering contributes to maintaining the systems of concepts, language, and social norms that underpin the aforementioned problems. While public discourse significantly influences socially operative concepts, private discourse also reinforces ideology (Dembroff & Wodak, 2018). Among these reasons, I focus particularly on point (i), i.e., the argument concerning the expressions of disrespect and the psychological consequences of misgendering.

Research consistently shows that microaggressions negatively affect the mental health of members of marginalised social groups. They are associated with lower self-esteem, higher prevalence of depressive symptoms, lower levels of psychological well-being, and greater intensity of negative emotions (Nadal et al. 2016). Nadal et al. (2014) highlighted common reactions to microaggressions experienced by trans* individuals. The authors identified three response domains to microaggressions: emotional, behavioural, and cognitive (Nadal et al., 2014). Firstly, (i) the emotional response domain includes feelings of anger, betrayal, despair, exhaustion, and feeling invalidated and misunderstood. Secondly, (ii) the behavioural scope of the response domain includes confronting the microaggressor by setting boundaries or avoiding situations where microaggressions are likely to occur. Other behavioural responses include passive coping strategies,[10] such as avoidance, or attempting to appease the microaggressor in order to defuse a tense situation. Thirdly, (iii) cognitive reactions include rationalisation, such as seeking justification by assuming that the microaggressor is uneducated. Consistent emotional reactions involving negative

10 The American Psychological Association defines a coping strategy as "an action, a series of actions, or a thought process used in meeting a stressful or unpleasant situation or in modifying one's reaction to such a situation". Further information can be found at: https://dictionary.apa.org/coping-strategy.

feelings and emotional distress have been identified, as well as similar cognitive strategies such as rationalisation and behavioural strategies to avoid or minimise confrontation (Nadal et al., 2014).

Research shows that the psychological consequences of misgendering are consistent with those of identified in cases of microaggressions in general. Kevin A. McLemor (2018) reports that misgendering is associated with psychological distress and the perception of the presence of stigma in society. Having a social support network helps individuals to identify instances of prejudice and discrimination more accurately. Consequently, individuals who report high levels of social support tend to perceive a stronger correlation between the frequency of misgendering and the feeling that trans* individuals are more widely stigmatised in society (McLemore, 2018).

However, the impact of misgendering can vary depending on several factors, such as the context in which it occurs and how much a person's sense of self is defined by their collective identity. It can also depend on how it intersects with other forms of discrimination, such as those based on race, class, or sexual orientation. The intent behind the misgendering is also relevant: intentional misgendering tends to be perceived as more harmful than unintentional errors, the distress caused by which may vary depending on the circumstances (Gibbs, 2024).

In their work on the effects of microaggressions on gender non-conforming individuals, Arijs et al. (2023) report that being referred to using incorrect pronouns and names is commonly experienced, particularly within the family context. These incidents often arise from a lack of acceptance or entrenched language habits. Participants described a range of reactions: some felt that being misgendered by relatives amounted to a denial of their identity, leading to feelings of gender dysphoria; others interpreted the misgendering as unintentional and attributed it to unconscious habits rather than hostility. Nevertheless, many participants reported that repeated misgendering caused emotional distress, self-doubt, and reluctance to express their gender identity openly. Some also felt inhibited in their identity development or even began to question the validity of their own gender identity. As one interviewee explained: "It had the effect on me that I stopped myself to be open about my gender identity to my environment, and to myself" (Arijs et al., 2023, p. 15). Another admitted: "First I thought, am I just kidding myself, like am I being serious with myself but yeah, I know that I am still searching so then I sometimes got remarks like oh you're just confused, you know" (Arijs et al., 2023, p. 15).

In conclusion, research on misgendering highlights the negative psychological impact frequently experienced by trans* individuals, effects that are broadly consistent with findings from research on microaggressions in general, which have been discussed previously (§ 3.3). Overall, these harms significantly affect mental health, and their severity varies depending on the aggressor's intent, the social context of the interaction, and the victim's perception of the act. As Dembroff and Wodak (2018) argue, not all four reasons (i-iv) need to apply equally, or even apply at all, in every instance of misgendering to support the moderate claim. In general, these arguments suffice to establish the moral wrongness of misgendering, even if only some reasons are present or vary in intensity depending on the context. Their argument holds as long as enough of these reasons are in play to ground a duty not to misgender, in both binary and non-binary trans* cases (Dembroff & Wodak, 2018). The psychological and medical literature reviewed here supports this claim, as it shows that misgendering has clear negative psychological consequences for trans* individuals. These harms affect people's lived experiences and sense of identity, making gendered language a salient and potentially violent speech act when used in a way that disrespects someone's self-identification.

4. The duty not to use gender-specific pronouns altogether

As anticipated, Dembroff and Wodak (2018) also discuss a second claim, the radical one. The radical claim asserts that we have a moral duty not to use gender-specific pronouns altogether, including when referring to binary individuals. This means avoiding both "he" and "she", regardless of the referent's gender identity, and also rejects non-binary pronouns such as singular "they". To support this claim, the authors present three arguments.

The first argument (i) is that there is no better alternative than eliminating gender-specific pronouns altogether. According to Dembroff and Wodak (2018), once we accept the moderate claim, that we have a moral duty to avoid using binary pronouns for non-binary individuals, two main options remain if binary pronouns ("he" and "she") are to be retained:

(1) Introducing a third, catch-all pronoun (e.g., singular "they" or "ze") for non-binary referents.

(2) Using a distinct pronoun for each non-binary identity.

Both approaches are considered problematic. The first (1) is inegalitarian: it preserves two specific pronouns ("he" and "she") for men and women while collapsing all non-binary diversity into a single generic pronoun (singular "they" or "ze"). This reinforces the perception that binary identities are more "natural" or legitimate. The second option (2) is deemed unfeasible due to the cognitive burden of learning and using an open-ended set of new pronouns, especially given that pronouns are a closed grammatical class and non-binary identities are fluid and evolving. For these reasons, the authors conclude that eliminating gender-specific pronouns altogether offers the most consistent and equitable solution (Dembroff & Wodak, 2018).

The second argument (ii) that Dembroff and Wodak (2018) discuss in support of the radical claim is that providing gender information can compromise the referent's privacy. They argue that the use of gender-specific pronouns (such as "he" and "she") often places individuals in a morally problematic position they call "Disclose or Deceive": they must either disclose private information about their gender identity or sexual orientation, or deceive others, explicitly or implicitly, about these aspects of themselves. Through illustrative examples, they show how individuals like Asher (a trans man) or John (a gay man) may be forced to correct others' assumptions to avoid misgendering or heteronormative assumptions, thereby revealing sensitive information they may prefer to keep private. Alternatively, by remaining silent, they risk misleading others and potentially being perceived as dishonest if later "outed" (Dembroff & Wodak, 2018).[11]

Finally, the third argument (iii) advanced by the authors to support the radical claim concerns the harmful cognitive and social effects associated with the use of gender-specific pronouns and other gendered linguistic markers. Dembroff and Wodak (2018) argue that linguistic devices such as pronouns, gendered suffixes (e.g., "-ess" in hostess), and grammatical gender in general do not merely reflect but actively shape social beliefs about gender. In particular, they claim that gendered language contributes to the reinforcement of essentialist

11 The authors further argue that even adopting gender-neutral pronouns (like singular "they" or "ze") does not fully solve the issue, as these forms can carry pragmatic implications, for instance, suggesting that someone is non-binary or hiding something, thus still putting pressure on individuals to disclose. Importantly, they reject the idea that this is merely a symptom of broader discrimination: gender-specific pronouns themselves exacerbate the problem by embedding assumptions about gender and making them harder to challenge. Even in a world without discrimination, they argue, individuals have a legitimate autonomy-based interest in controlling when and how to disclose intimate aspects of their identities. Ultimately, they contend that because pronouns are pervasive and presuppose gender information, they systematically contribute to the reproduction of discriminatory norms and thus represent not just a symptom but part of the cause (Dembroff & Wodak, 2018).

beliefs, i.e., the idea that gender is an intrinsic and explanatory feature of individuals, shaping psychological traits and social roles (Dembroff & Wodak, 2018).[12]

Although the use of gendered language can reinforce essentialist assumptions, it is important to acknowledge the positive and empowering roles that gendered linguistic resources can play, particularly for trans* individuals. Pronouns and proper names are fundamental to affirming gender identity and social inclusion (Gibbs, 2024). Far from being mere expressions of a binary system, these linguistic elements are often actively used by trans* people to assert agency, affirm identity, and foster joy. In this regard, E. M. Hernandez and Archie Crowley (2024) have criticised the tendency, common in philosophical debates on gender and language, to focus primarily on the potential harms of gendered languages while overlooking how trans* individuals themselves use such language in constructive and life-affirming ways. According to the authors, trans* people engage with language through what they call "recognition use", "playful use", and "joyful use", three modes of linguistic agency that contribute meaningfully to personal and collective well-being (Hernandez & Crowley, 2024).

The recognition use involves employing gendered linguistic resources, such as pronouns, labels, and names, to accurately represent one's gender identity, enabling one to be recognised as such by others. For example, in societies with binary gender norms, creating and using non-binary labels enables individuals to express their gender identity and be understood, thereby fostering both self-definition and mutual understanding (Hernandez & Crowley, 2024).[13]

The playful use, drawing on Maria Lugones's (2003) notion of "playfulness", refers to the subversive appropriation of gendered language that resists oppressive norms without fully internalising them (Lugones, 2003). Hernandez and Crowley (2024) give the example of trans* individuals who use gendered anatomical terms in non-standard ways, for instance, to refer to their bodies regardless of surgical status. One example is trans* people who refer to their genitals as "dicks" or "cocks" without having undergone genital surgery, thereby reclaiming and re-signifying these terms in a way that affirms their gender identity without conforming to cisnormative expectations (Zimman, 2014). This recontextualisation allows them to explore gendered language creatively, without adhering to its normative "rules", and to engage critically with dominant gender ideologies. For instance, by using terms like "dick" to refer to a vulva or by reclaiming traditionally gendered anatomical labels in ways that affirm their own embodied experiences, trans* individuals unsettle the assumed correspondence between language, anatomy, and identity, revealing the contingency and normativity of dominant linguistic conventions (Hernandez & Crowley, 2024).

Finally, the joyful use refers to the profound sense of euphoria that many trans* people report when addressed with gender-affirming names and pronouns. Studies have shown that the use of appropriate gendered language is closely linked to psychological well-being and

12 To illustrate this point, the authors present a thought experiment: imagine a fictional world where people use different pronouns depending on body weight, "fee" for thin individuals and "fum" for fat individuals. In such a world, even those who usually accept gender-specific pronouns would likely find this linguistic system objectionable, since it implies that body weight is always relevant to communication. The example demonstrates how pronoun systems can make certain traits permanently salient, reinforcing their perceived importance in all social contexts. Dembroff and Wodak (2018) argue that gender-specific pronouns in our world work similarly: by routinely invoking gender, they convey the implicature that gender is a meaningful and explanatory feature of people's behaviour and social identity, even when it is not (Dembroff & Wodak, 2018).

13 Hernandez and Crowley (2024) present a series of pertinent examples that show how diverse linguistic communities use gendered linguistic resources to recognise varied identities, even in the absence of gender pronouns. In some cases, these communities use these linguistic resources to challenge traditional gender expectations. For a more comprehensive overview, please refer to Hernandez and Crowley (2024).

gender euphoria (Beischel et al., 2022; Hernandez 2021). As Hernandez and Crowley (2024) observe, this dimension of language use highlights its capacity not only to validate but to generate joy and self-affirmation. Consequently, as the authors observe, the use of gender-appropriate language in expressing one's gender can elicit profound feelings of joy. This joy is not merely affective but politically significant: it emerges as a form of embodied affirmation that resists cisnormative structures and affirms the legitimacy of one's self-identification. In this sense, joy becomes a transformative effect that challenges the marginalisation of non-normative identities and asserts the right to linguistic and existential self-determination (Hernandez & Crowley, 2024).

Taken together, these forms of engagement show that gendered language is not inherently oppressive or reductive, but rather a flexible and powerful tool. Trans* individuals use it in ways that range from affirming identity to critiquing and reimagining normative structures. This underscores the significance of pronoun usage and the linguistic resources available for effective gender identity expression. Furthermore, Hernandez and Crowley (2024) contend that the profound applications of gendered language can prompt a re-evaluation of Dembroff and Wodak's (2018) radical claim, potentially rendering it subject to criticism. The project of eliminating any personal reference to gender may not be the optimal approach, as it would result in the loss of the full range of benefits that gendered language offers (Hernandez & Archie Crowley, 2024).

Finally, as Dembroff and Wodak (2018) note, their argument is developed with English in mind, a language that has gender-specific pronouns but does not explicitly encode gender pervasively. They acknowledge that their radical proposal might be easier to implement in English than in languages with more pervasive gender systems, such as Portuguese, and leave open whether it could apply to them (Dembroff & Wodak, 2018).[14] As Francesca Cesarano (2022) observes, in grammatical gender languages, even if gender-specific pronouns were to be avoided, which is a challenging enough task in itself, gender information would still be pervasively provided in most nouns and target agreement. Therefore, while it may be possible to avoid gender-specific pronouns in languages such as English, this could be more difficult in grammatical gender languages (Cesarano, 2022). However, this topic will not be explored further in this work due to space limitations.[15]

Conclusion

Social recognition of gender identity is fundamental to both individual well-being and the construction of a personal gender identity. Whether intentional or not, misgendering is a form of microaggression that has the potential to cause significant psychological distress and

14 English, according to the classification of Stahlberg et al. (2007), is a "natural gender language". In natural gender languages, such as English, gender information is not provided as pervasively as it is in other languages. Gender information is mainly provided by gender-specific pronouns such as "she" and "he" and by lexically gender-marked nouns such as "mother" and "father". In other languages, called "grammatical gender languages", such as Italian, French and Spanish, it is very difficult to avoid giving gender information. Grammatical gender languages have a grammatical gender system, and most nouns are assigned a gender. In the case of animate entities, grammatical gender typically aligns with the referent's gender (Corbett, 1991; Thornton, 2022). For human referents, terms are declined in the masculine or feminine form based on referent gender, such as the Italian personal nouns "maestra" (teacher[F]) and "maestro" (teacher[M]), which are typically assigned to a feminine and masculine referent, respectively. Furthermore, personal pronouns and other grammatical forms align with the gender of the noun, thus also conveying information about the gender of the person to whom the noun refers. Therefore, in these languages, gender information is provided pervasively (Stahlberg et al., 2007).

15 In her exploration of the phenomenon of misgendering, Francesca Cesarano (2022) examines the discourse surrounding the moderate and radical claims of Dembroff and Wodak (2018). Through discussing these claims, Cesarano (2022) also reflects on the feasibility of implementing the project in languages such as Italian, highlighting the associated challenges. For a more comprehensive overview, please refer to Cesarano's work (2022).

to reinforce systemic exclusion. As I have discussed in this work, misgendering has an impact that extends beyond interpersonal interactions, shaping broader social and institutional dynamics.

Having examined the philosophical arguments presented by Dembroff and Wodak (2018), I conclude that their moderate claim that we have a duty to avoid misgendering both binary and non-binary trans* individuals is strongly justified. This conclusion is supported by compelling arguments, particularly with regard to the psychological harm caused by misgendering and its role in perpetuating systemic injustices. However, I find the authors' radical claim that we should eliminate all gender-specific pronouns to be more contentious. While I acknowledge the potential benefits of such a shift, I conclude that gendered language plays an important role in constructing personal identity and affirming gender.

Particularly interesting is what Dembroff (2018) emphasises the importance to observe trans* and queer contexts for understanding how linguistic and social practices can be reconfigured to better support gender diversity. In particular, many trans* communities have developed gender-neutral linguistic strategies, designated inclusive physical spaces, and practices centred on the use of self-chosen names and pronouns. These structures foster greater autonomy in gender expression and challenge normative gender roles, thereby creating social environments in which individuals can affirm and reclaim their identities (Dembroff, 2018).

Furthermore, given that misgendering often occurs unintentionally, it is important to reflect on the reasons behind these instances and adopt strategies to counteract them. As previously discussed (§ 2), people may unintentionally commit misgendering for various reasons, such as misunderstanding, misinterpreting gender identities, or lacking sufficient information about gender (Nordmarken, 2014). This underscores the importance of increasing awareness about gender issues and respecting gender identities by using appropriate language. Institutions can play a crucial role in addressing the lack of awareness about gender diversity and the importance of mindful communication through education. However, delving into institutional strategies to counter misgendering is beyond the scope of this paper.

It is worth noting that much of the existing psychological and medical literature focuses on cases of misgendering involving binary trans* individuals, with trans* individuals who identify as binary, with individuals who identify as non-binary being absent from the discussion. In other cases, researchers do not distinguish between binary and non-binary trans* instances. This implies that the differences in misgendering experiences between these two categories are not adequately captured (Jacobsen et al., 2023). Fortunately, however, attention to this topic beyond gender binarism is growing in literature. With this article, I have aimed to explore the psychological implications of misgendering and highlight why this type of harm can be used to argue against misgendering. These considerations lay the groundwork for addressing other issues, such as the differences in experience between binary and non-binary trans* individuals. Understanding these differences could help to identify ways to counteract this type of microaggression and, more broadly, help to contrast the perpetuation of ethnocentric values, prejudices, assumptions, and stereotypes.

REFERENCES

Arijs, Q., Burgwal, A., Van Wiele, J., & Motmans, J. (2023). The Price to Pay for Being Yourself: Experiences of Microaggressions among Non-Binary and Genderqueer (NBGQ). *Youth. Healthcare*, *11*, 742. https://doi.org/10.3390/healthcare11050742;

Beischel, W. J., & Van Anders, S. M. (2022). A little shiny gender breakthrough: Community understandings of gender euphoria. *International Journal of Transgender Health, 23*(1), 1-15. https://doi.org/10.1080/26895269.2021.1915223;

Cesarano, F. (2022). On misgendering and the feasibility of changing grammar rules to respect all gender identities. *Degree Thesis, Scuola Universitaria Superiore IUSS*;
Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press;
Dembroff, R., & Wodak, D. (2018). He/She/They/Ze. *Ergo, an Open Access Journal of Philosophy 5* (20201214). https://doi.org/10.3998/ergo.12405314.0005.014;
Dembroff, R. (2018). Real talk on the metaphysics of gender. *Philosophical Topics, 46*(2), 21-50. https://doi.org/10.5840/philtopics201846213;
Gibbs, M. (2024). Misgendering as a Microaggression: gender Expression and Trans Pedagogy. *McNair Research Journal SJSU*, Vol. 20, Article 6;
Hernandez, E. M. (2021). Gender-Affirmation and Loving Attention. *Hypatia*, 36(4), 619-635. https://doi.org/10.1017/hyp.2021.45;
Hernandez, E. M., & Crowley, A. (2024). How to Do Things with Gendered Words. In E. Lepore & L. Anderson (Eds.), *The Oxford Handbook of Applied Philosophy of Language*. New York, NY: Oxford University Press;
Human Rights Campaign. (s.d.). Sexual Orientation and Gender Identity Definitions. HRC. https://www.hrc.org/resources/sexual-orientation-and-gender-identity-terminology-and-definitions;
Jacobsen, K., Davis, C. E., Burchell, D., Rutherford, L., Lachowsky, N., Bauer, G., & Scheim, A. (2023). Misgendering and the health and well-being of nonbinary people in Canada. *International Journal of Transgender Health*, 1-15. https://doi.org/10.1080/26895269.2023.2278064;
Jenkins, K. (2023). *Ontology and oppression: Race, gender, and social reality*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780197666777.001.0001;
Kapusta, S. (2016). Misgendering and its Moral Contestability. *Hypatia* 31 (3): 512-519;
Lugones, M. (2003). *Pilgrimages/Peregrinajes: Theorizing coalition against multiple oppressions*. Lanham, MD: Rowman & Littlefield;
Matsuno, E., & Budge, S. L. (2017). Non-binary/Genderqueer Identities: A Critical Review of the Literature. *Current Sexual Health Reports*, 9(3), 116-120. https://doi.org/10.1007/s11930-017-0111-8;
McLemore, K. A. (2018). A Minority Stress Perspective on Transgender Individuals' Experiences With Misgendering. *Stigma and Health, 3*(1), 53-64. https://doi.org/10.1037/sah0000070;
Nadal, K. L., Whitman, C. N., Davis, L. S., Erazo, T., & Davidoff, K. C. (2016). Microaggressions Toward Lesbian, Gay, Bisexual, Transgender, Queer, and Genderqueer People: A Review of the Literature. *The Journal of Sex Research*, May-Jun, 53(4-5): 488-508. doi: 10.1080/00224499.2016.1142495;
Nadal, K. L., Davidoff, K. C., Davis, L. S., & Wong, Y. (2014). Emotional, behavioral, and cognitive reactions to microaggressions: Transgender perspectives. *Psychology of Sexual Orientation and Gender Diversity*, 1(1), 72-81. https://doi.org/10.1037/sgd0000011;
National LGBTQIA+ Health Education Center. (2024). *LGBTQIA+ Glossary of Terms for Health Care Teams*. https://www.lgbtqiahealtheducation.org/publication/glossary/;
Nordmarken, S., & Kelly, R. (2014). Limiting transgender health: Administrative violence and microaggressions. In V. L. Harvey & T. H. Housel (Eds.), *Health care disparities and the LGBT population*. Plymouth: Lexington Books;
Nordmarken, S. (2014). Microaggressions. *TSQ* 1, 1 (1-2): 129-134. https://doi.org/10.1215/23289252-2399812;
Rogers, M. M. (2021). Exploring the Domestic Abuse Narratives of Trans and Nonbinary People and the Role of Cisgenderism in Identity Abuse, Misgendering, and Pathologizing. *Violence against Women*, 27 (12-13): 2187-2207. https://doi.org/10.1177/1077801220971368;
Sevelius, J. M. (2013). Gender Affirmation: A Framework for Conceptualizing Risk Behavior

among Transgender Women of Color. *Sex Roles*. Jun 1; 68(11-12): 675-689. 10.1007/s11199-012-0216-5. PMID: 23729971; PMCID: PMC3667985;

Sevelius, J. M., Chakravarty, D., Dilworth, S. E., Rebchook, G., & Neilands, T. B. (2020). Gender Affirmation through Correct Pronoun Usage: Development and Validation of the Transgender Women's Importance of Pronouns (TW-IP) Scale. *International Journal of Environmental Research and Public Health*, Dec 19; 17(24): 9525. 10.3390/ijerph17249525. PMID: 33352630; PMCID: PMC7766835;

Sue, D. W. (2010). *Microaggressions and marginality: Manifestation, dynamics, and impact.* Hoboken, NJ: Wiley & Sons;

Thornton, A. M. (2022). Genere e igiene verbale: L'uso di forme con ə in italiano. *Annali del Dipartimento di Studi Letterari, Linguistici e Comparati. Sezione linguistica*, 11-54 Paginazione. https://doi.org/10.6093/2281-6585/9623;

Zimman, L. (2014). The discursive construction of sex: Remaking and reclaiming the gendered body in talk about genitals among trans men. In L. Zimman, J. L. Davis, & J. Raclaw (Eds.), *Queer Excursions: Retheorizing Binaries in Language, Gender, and Sexuality* (pp. 13-34). https://www.researchgate.net/publication/299892505_The_Discursive_Construction_of_Sex.

SECTION 4

SECTION 4

ART, AESTHETIC EXPERIENCE, AND THE PERFORMANCE OF SELFHOOD

CHIARA CAPPELLETTO
Università degli Studi di Milano
chiara.cappelletto@unimi.it

WHY BRAIN IMAGES ARE NOT REPRESENTATIONS OF THE SELF: THE TRANSPARENCY OF THE BRAIN/BODY/MACHINE SYSTEM[1]

abstract

Brain imagery is central to the representation of personal identity at the expenses of the living body. This essay explores how visual and performance art, cognitive science, and neuroscience have contributed to the dissemination of the idea that to "be" means to "have a brain". Brain scans as forms of "inner portraits" have naturalized a reductive vision of personhood, overshadowing the entanglement of the body with the environment and technological apparatuses. The supposed transparency of brain images is questioned through artworks by Abdoulaye Konaté, Jan Fabre, Pierre Huyghe, and Refik Anadol among others, and discussed via the concept of visibilization: a medial process that renders visible what is not visual. MRI and fMRI are shown to be a acoustic operation involving the brain, the body, imagers, and the environment in a relational system. The article calls for a rethinking of human subjectivity as part of material, epistemic, and narrative networks.

keywords

brain images, brainhood, media transparency, personal identity, entanglement, visibilization

1 I thank Giulio Galimberti and Paola Zilioli for their valuable comments and Samuel Fleck for his careful reading of the text.

Phenomenology and Mind, n. 29 - 2025, pp. 140-150
DOI: 10.17454/pam-2912
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Public Brains, Neglected Bodies and Invisible Arguments

For over fifty years, visual and performance art have depicted "the brain" as a metonym for "personal identity," displaying it in a wide range of media and settings. The introduction of the cerebral organ into the domain of the visible as a general portrait of human beings induces a neglect of their phenomenological bodies. As Paul Ricoeur sharply observed (1990), "the cerebral fictions used to discuss personal identity neutralize the body and restrict it to the brain, at the expense of the self as flesh" (p. 378). This narrative originated in the West. In fact, "no other culture has proposed the reducibility of Self to an organ of the body" (Vidal, 2009, p. 11) and built a whole rhetoric around it. That notwithstanding, the embrained notion of the Self has by now been registered in countries and contexts all over the world. Thus exported, the visual telling of a global cerebral narrative continues to unfold, all too often without further explanation of how brains, bodies, and personhood are related, and what part images play, together with the devices needed to produce them. All in all, brain scans are highly effective vehicles for defining a new, globally accepted concept of personal identity.

Just consider the work *Open Mind* by Cuban artist Yoan Capote, featured at the 10th Havana Biennial (2009). A brain-shaped labyrinth with walls made of painted PVC and metal replicating the cerebral hemispheres and neural folds invited visitors to walk through as if they were neurons traversing thought processes. The installation was site specific and set in dialogue with the Morro Castle, suggesting a relation between historical memory and personal consciousness. Extracted from the skull, the brain had materialized as a standalone object. The Malian artist Abdoulaye Konaté, a leading figure in African contemporary art, in 2016 created one of his large-scale, colorful textile works, titled *Non au fanatisme religieux*. It bears a transcultural message of secularism, rendered in French yet oriented from right to left, following the direction of the Arabic script. Embroidered in white at its center is a brain. It serves as a symbol of rational thought, based on an idea of the Self as discrete and independent of beliefs and opinions pertaining to the sensible world.

This epistemic account, which characterizes Western modernity, has flourished since its inception. Gilbert Ryle (1949) has shrewdly observed that the ideology of the Self as a "pure" cognitive agent stems from Descartes (and more notably, his readership), who conceived of human beings as articulated in an "inside" where cognition happens privately and in the same way for everybody and an "outside" where social relations happen thanks to specific bodily behaviors occurring in cultural environments. This divide between the secret mind and the public body has survived for centuries, notwithstanding critiques leveled at it by

philosophers—notably contemporary ones (Deleuze, 2024; Braidotti, 2002)—as well as by some neuroscientists (Damasio, 1994; Gallese & Morelli, 2024). However, the knowledge that cognition depends on neurobiological mechanisms complicates the split, and even contradicts it, because in vivo individual brain activity enters the public eye and does so through scans. We can finally see our minds pictured, can't we?

The general understanding is that brain images, once outside the medical context, can be looked at the same way as portraits are (whether painted, photographic, video, or otherwise generated). The onlooker's behavior is analogous in both cases, for the greater glory of the brain that I am as a person. A striking overlapping is at stake between the individual, the cerebral matter, the image, and the medial and conceptual tools (the technology and the knowledge) required to produce my brain scan. For, brain imaging aligns with "a philosophically informed methodological individualism, in the sense that for many theorists the goal is to identify mechanisms found within an individual, and specifically within an individual brain, to explain how that individual understands other people" (Gallagher, 2020, p. 3). Along the same lines, Fernando Vidal (2009) argues that

> the individualism characteristic of western and westernized societies, the supreme value given to the individual as autonomous agent of choice and initiative, and the corresponding emphasis on interiority at the expense of social bonds and contexts, are sustained by the brainhood ideology and reproduced by neurocultural discourses. (Ehrenberg, 2008, p. 7)

Over the past fifty years, a notion of human beings that is deeply shaped by medial practices grounded in brain activity has emerged.

> To have the same brain is to be the same person, and [...] the brain is the only part of the body we need in order to be ourselves. As a 'cerebral subject,' the human being is specified by the property of 'brainhood,' i.e. the property or quality of being, rather than simply having, a brain. (Vidal, 2009, p. 6. See also Ferret, 1993, p. 79)

An interesting instance of the alliance between our old trust in portraits, our new reliance on cerebral scans to grasp the inner Self, and the cerebral subject emerges from a conversation between Bruno Latour and Paul Churchland. The former, a leading figure in science and technology studies, reports the exchange he had with the latter, a prominent neurophilosopher, according to which common sense psychology provides radically false theories on human cognition and should be completely replaced by neuroscience (Churchland, 1984). Despite his theoretical normativity, Churchland was carrying in his wallet a color scan of his wife's brain, just as one might carry the photograph of a loved one taken in a photo booth. According to Latour (2004),

> Paul insist[ed] adamantly that in a few years we will all be recognizing the inner shapes of the brain structure with a more loving gaze than noses, skins and eyes! Unquestionably, Paul sides with the eliminativists: once we have a way of grasping the primary qualities (in his case the brain macrostructure, but it could be, for other even more advanced scientists, the microstructures of individual neurons, or the DNA sequences of the brain itself, or even further, the atomic structure of the biophysics of the DNA [...]), we can eliminate as irrelevant all the other versions of what it is to be a body, that is, to be somebody. (Latour, 2004, p. 224)

That Churchland's wife, Patricia, may not have objected to her large blue eyes, blond hair, and white skin being replaced by a cerebral scan conventionally colored is irrelevant to the matter at hand. Her belief that "I am who I am because my brain is what it is" (Churchland, 2013, p. 11) only underscores the extent to which the medial character of the cerebral data that machines extract, collect, and make available to a large audience through images is disregarded in ways that neglect the lived and living body: in this instance, hers.

Even more, the embodied nature of the brain and the embedded nature of the brain/body system are bracketed. In the case of brain images serving as representations of oneself, the bracketing occurs because they are conceived of as transparent means of accessing the brain as the primum mobile of personal identity. It's right there for us to see, as if we had just opened a window in the skull. The "inside" has then been externalized, pace Descartes, and our brains have become "sites where organs and eyes meet", to paraphrase Michel Foucault (1963. See also van Dijck, 2005). Such a "meeting" is possible thanks to the idea that imaging techniques do not affect our "looking at", as if they were imperceptible.

An interface that "erases itself, so that the user is no longer aware of confronting a medium, but instead stands in an immediate relationship to the content of that medium" is called "transparent" (Bolter & Grusin, 1999, p. 24). That "leads one either to erase or to render automatic the act of representation" (ivi, p. 33), that is immaterial and impersonal. This is what happens to brain imaging when scans leave the medical room and enter the layperson world. The "scanned" brain—whether printed or on-screen—suggests, on the one hand, a denial of the machinery at work and, on the other, an obliteration of the human body under the scan, as if beds, computers, screens, cables, radiologists did not enter in the image making process, and a specific human being with all his/her physical idiosyncrasies was not laying there. Such a neglect of the personal body in its situatedness and material entanglement goes hand in hand with medial transparency and leaves us in a blind spot.

But for transparency to occur, not only must the material and medial conditions of images themselves be neglected; the knowledge system that configures our understanding of ourselves and others—and dictates how we act in accordance with it, in a loop of retroaction and mutual shaping—must also be overlooked, since no medium develops in isolation from the theoretical positions under which it is experienced, acknowledged and interpreted; each medium works also as a network of technical, political, and economic forces, which are an integral part of specific ways of thinking. The invisibilization of the brain imaging techniques is then theoretical, medial, and material and affects scholarly accounts. The obliteration of medial and intellectual entanglement induces an epistemic blindness with respect to the way in which we explain ourselves. It deeply affects us as epistemic agents, so much so that even a rigorous eliminativist scholar believes in the power of some kind of images to make the beloved present.

The self-erasure of media ally with the self-erasure of the genealogy and the implication of theoretical arguments that stem from the common sense that they nourish in turn (Wittgenstein, 1972; 1989). Notably, the actual material condition under which the intellectual activity is carried out is overlooked, as if we were purely abstract thinkers (Ahmed, 2006). Feminist philosophy has been effective in unveiling the extent to which the supposed clarity of theoretical propositions and their subsequent efficacy depends from the disregarding of their own situatedness, despite the intertwining of first- and third-person perspectives and the tools and devices used to direct our understanding. The competition among conceptual accounts arising from such a disciplinary narrowing leads to specialized claims and a fragmented and scattered understanding of the Self.

To mitigate the paradox of a quest to know who we are that ultimately obscures how we live and think, many philosophers in recent decades have argued for a multidisciplinary

approach that addresses personhood in its full complexity. Shaun Gallagher (2013) proposes a "Pattern Theory of Self" to advocate for a multilevel perspective to understanding human nature. He posits that "to stay plural about the concept of the self," (1) we should not consider physical, social, private aspects as mere modifiers of an independently existing Self, and invites us to think of them "as organized in certain patterns, [...] a particular variation of [which] constitutes what we call a Self" (1). According to him, the Self results from various contributory elements, "none of which on their own is necessary or essential to any particular self. This is not a pattern theory of 'the Self.' Rather, what we call 'self' is a cluster concept that includes a sufficient number of characteristic features" (3). He identifies several groups of aspects encompassing neurobiological, affective, behavioral, interpersonal, cultural, material, and narrative dimensions. Gallagher's proposal has the advantage of transcending any nature/culture or mind/body dualism and accommodating different types of experiences, knowledge, and media. More important, it conceptualizes the Self as an agent capable of acting and being acted upon, rather than a detached and encapsulated observer of the surrounding world. However, the dynamics of cooperation or conflict at work in the (cognitive and corporeal) Self remain to be fully acknowledged and described.

As often happens in the cognitive sciences, the discourse blurs when addressing the actual conditions that shape the formation, expression, and recognition of the Self in the mundane world and how they sensibly intra-act (Barad, 2007) with our bodies to different degrees and in different ways. Cognitive neuroscientific understanding, based on brain visualization, requires then a more fitting acknowledgment of how such techniques feed back into the brain/body system. Interestingly enough, when such an enterprise is undertaken, transparency is not questioned. For instance, Gallagher's critique, whereby "in an fMRI experiment [...] the experimental subject is necessarily alone in a large noisy machine, where bodily movement is extremely limited" and that impedes his/her study, since "our primary form of social cognition depends at all on bodily, in-person interactions with other people, rather than passive observations [...] (see Schilbach et al., 2013; Gallagher et al., 2013)" (2020, p. 3), does not challenge the image making process.

In the end, the powerful visual narrative in which brain images have trapped us affects the whole epistemic enterprise whereby we understand our identities, and ultimately shoddily equates humans with internal organs. I will try to pick apart this narrative by articulating the entanglement of living matter and machines, of brains and bodies, of human beings and the environment they inhabit and die in, showing the interplay between narrative props, our living cerebral organ, and its images.

2. Embedded Machines

Self-narration has always produced an extraordinary quantity of personal images (Giannachi, 2023). The history of portraiture—from the legendary figure of Butade's daughter to the Montmartre painters—and self-portraiture—from Albrecht Dürer to Bruce Nauman and present-day selfies—offers substantial evidence of this phenomenon, which occurs in different techniques, styles, and time periods, yet consistently privileges the surface of the frontal part of the head. Portraiture as a visual technology of the Self reinforces the idea that personhood is discrete and localized within a framed space, and relegates its surroundings to the backgrounds. This is rooted in a physical condition, which is the insularity of the face with respect to the rest of human anatomy (Simmel, 1901; Deleuze & Guattari, 1980). The face is "beheaded," cut out from the rest of the body, at the cost of freezing our living personhood (Cappelletto, 2025).

The downplaying of the body in favor of one of its most prominent part is redoubled by brain images, which have permeated popular culture since the eighties, even before the neuroscientific hype of the nineties, which then-President George Bush referred to as the

decade of the brain. The July 1983 issue of Vogue featured an article titled "New Seeing-Eye Machines," presenting three vividly colored PET (Positron Emission Tomography) scans accompanied by a caption for a lay audience: "These brain scans show—graphically—how a normal brain's function compares to that of a depressed or schizophrenic patient" (Hixson, 1983). As the anthropologist Joseph Dumit (2003) has remarked: "These images persuade viewers to equate person with brain, brain with scan, and scan with [medical] diagnosis" (p. 36), which offers a particular type of narrative. Notably, machines are equated with eyes. In so doing, their medial and material character is erased, despite the costs, energy, scientific knowledge, and technical know-how associated with operating them.

Twenty years after that first publication, in 2003, the conceptual artist Jonathon Keats underwent functional magnetic resonance imaging (fMRI) while contemplating themes like art, beauty, love, and death. The resulting images were exhibited at the San Francisco Modernism Gallery, where they served a novel purpose: to sell his brain activity, which he identified as an artwork generated by his cerebral processes, in a unbroken continuum. No methodological perplexity was raised by the idea that Keats's brain could be treated as independent of Keats the person, or that Keats's blue, red, white, and black scans were not immediately linked with his brain activity. Moreover, while the 1983 Vogue article tried to verbally describe as accurately as possible the scans' production and the role played by imagers and radiologists, subsequent uses of brain imagery in nonscientific or nonmedical contexts have largely dispensed with explanations.

Mesmerized by its graphics, we witness the old story of the inner Self now mechanically transduced into the public eye to be seen, grasped, and understood. As Jean Baudrillard (1986) anticipated, "all that fascinates us is the spectacle of the brain and its workings," and "what we are wanting here is to see our thoughts unfolding before us" as if they were born out of Zeus's head. A strong bond has developed "between image making and the scientific enterprise aimed at determining neural processes of subjectification: [...] It would seem that neuroscience has saturated the explanatory field of the Self, aided by depictions of cortices, thalami, and the like" (Cappelletto, 2022a, p. 1). In 2007 brain imaging had evolved to the point that the American Psychological Association (2007) saw fit to celebrate fMRI as a technique that "produces movies starring the brain", as if there were no mediation at stake.

The first artist to use MRI was Justine Cooper in 1998 with her video animation *RAPT I*, which attracted broad public attention. Since then, two main types of visual evidence testify to the spectacular interest in neurological mechanisms devoid of proper phenomenological appearance, which serves as raw material for artists to work with.

In the first type, the visual outputs are replicated with materials from craft and artistic traditions that long predate brain imagers. In 2011 Dario Ghibaudo wove prayer rugs dedicated to the seven deadly sins, each localized in a different, color-coded area of the brain. The idea is that neurology suggests that sins are part of human mental activity, irrespective of one's faith. Mixed techniques are also employed in the ambitious multimedia project led by Jan Fabre, which combines performance, video, photography, and drawing. In *Do We Feel with Our Brain and Think with Our Heart?* (2013), the Belgian artist recorded videos documenting his dialogue with Giacomo Rizzolatti, the neuroscientist from Parma whose research team discovered mirror neurons in the nineties. Both the artist and the neuroscientist wore EEG electrodes, as they would in the experimental setting, which was conceptually replicated. In 2017 Michela Martello, working at the threshold of fiber arts and the clinic, embroidered a sagittal cut of the brain in pink on an old piece of beige linen for her piece *Grandmas Mirror*. In 2018 Laura Jacobson created *Resonance Punctuated CXLIV* in painted ceramic and wood, producing a form that clearly recalls a horizontal scan, although the object looks like a handmade craft assembled from scrap materials. Brain images are also integrated in multimedia artworks. In

2025 Laia Abril hung brain images in her installation *On Mass Hysteria* to visualize emotions experienced by girls and women facing political or personal oppression, such as anger, sadness, or fear, which lead to fainting, tremors, and even laughter and trances.

In the second type of brain artwork, artists use the same kind of imagers as those employed in research or clinical settings. I will present three meaningful examples. The first is by Pierre Huyghe. The French artist has long explored the human/non-human relationship, more recently redefined as the living/non-living relation. In his *UUmwelt* series (2012) Huyghe asked participants to enter a fMRI scanner, where they "were given verbal descriptions or shown pictures of living creatures, prehistoric tools, machines, codes, and artworks; the data were then processed via an AI neural network [...] which learn[ed] and progressively reconstruct[ed] the images by matching them with pictures taken in the real world" (Cappelletto, 2022b, p. 93). *UUmwelt* was produced at the laboratory led by neuroscientist Yukiyasu Kamitani, who with Tomoyasu Horikawa in 2017 developed an fMRI method to show how brains that have seen or imagined pictures produce signals that can be reverted into digital images corresponding to those seen or imagined. They argue that "visual features extracted by computational models were successfully predicted from brain activity patterns" (Horikawa & Kamitani, 2017, p. 9). In 2014 the prolific artist and neuroscience researcher Suzanne Dikker collaborated on *Measuring the Magic of Mutual Gaze* with Marina Abramovic, who adapted it from her previous performance *The Artist Is Present*, staged at the MOMA in New York in 2010. The experiment-like setup called for two participants wearing EEG devices to engage in a mutual gaze for thirty minutes while their brain waves were recorded and visualized for the public. Synchronization of the brain waves would trigger a lightning animation, thematizing human connectedness. A similar setting, albeit with no artistic implications whatsoever, has been installed at the National Museum of Science and Technology in Stockholm. In *Melting Memories* (2018) Refik Anadol collaborated with the Neuroscape laboratory at the University of California, San Francisco, to collect EEG data from participants as they recalled long-term memories. A 32-channel Enobio device was used to record brain signals, with a focus on beta and theta waves, which are associated with short- and long-term memory, respectively. These data were then processed using machine-learning algorithms, including recurrent neural networks (EEGLearn), and transformed into three-dimensional visual representations. Notwithstanding the sophisticated machinery used, once displayed in galleries or museums, these works were presented as visual outputs. "Instead of critically probing the workings of brain imagers, artworks on brain images (brain art) glamorize them [...], confirming their representational and indexical efficacy, and do not actually challenge the technologies involved" (Cappelletto, 2022b, p. 92). Literalness prevails over artisticity, as the installation at the Swedish museum confirms.

Despite the magic of immediacy, however, the viewer is not obliged to succumb to the transparent visual rhetoric of those images. In fact, unlike portraits and regardless of their bold assertiveness, brain images do not justify self-mirroring as if they were facsimiles or snapshots of our cerebral condition; nor are they neuronal portraits or personal metonymies. They are rather a step in the process of constructing human beings as entangled in a brain/body/environment system where machines participate in a feedback loop through a multimodal sensible world. Precisely when we consider the second type of brain artworks and focus on the processes entailed in them, we realize that the relationship between the human brain and the brain scan is not vision-based.

Brain visualization techniques were not initially intended for the human brain only and may be applied to other living and nonliving matter, to reveal the composition of the particles without disassembling the main components. Paul Lauterbur, who pioneered MRI in the seventies, originally anticipated the technique finding many useful applications in studying

the "internal structures, states, and compositions of macroscopic and microscopic objects," including human tissues. He called it "zeugmatography," echoing the zeugma, a rhetorical figure that works by joining different parts of a sentence in a single element (Lauterbur, 1973). Such blending is what scanning is about.

Let's now describe how MRI works in humans. In our living body, hydrogen nuclei spins are normally random. When a person enters an MRI scanner, a strong magnetic field causes these spins to align either with or against the field (along the z-axis). A second, changing magnetic field (the x-y-axis) is applied via radiofrequency pulses, exciting the hydrogen nuclei and causing them to precess. When the pulse stops, the hydrogen nuclei relax back to alignment, releasing energy. This energy is detected as the MR signal by receiver coils. Additional gradient coils vary the magnetic field locally to determine the spatial origin of the signals. The analog MR signal is then digitized, and differences in intensity are turned into gray-scale pixels. Each pixel corresponds to a voxel, representing a small volume inside the patient's body. In sum, MRI images are computer-generated visual reconfigurations of physical data such as the relaxation times of hydrogen atoms that are found abundantly in the human body among others. They map the tissue-specific relaxation rates of the protons of water in different tissues (pathological tissues have different relaxation characteristics than healthy tissues, and this is why they are used within the medical context). Finally, the radiologist can modulate the shades of gray dynamically to locate the pathology in a process known as "windowing." Comparisons can also be made by changing the gray contrast in specific areas of the image; this process is known as "leveling."

> The MRI machine does not provide a transparent window into a reality that precedes the imaging process, but nor does it fabricate distinctions. It operates, rather, by instituting a mixed, cross-over environment, where technical and natural elements and forces come to intra-act and mutually influence each other in a law-like manner (Hoel, 2020, p. 310).

No skull has been open.

Despite the "look" MRI has, it is mostly a sound machine (Casini, 2011). The characteristic knocking or rocking sound heard during an MRI scan is caused by the activation of gradient pulses. In every MRI image, there is a combination of both signal and noise: the signal appears as the brightness of each voxel or pixel, while the noise is made up of random fluctuations in pixel intensity, often originating from the patient's body. This noise can interfere with the clarity of the image and needs to be minimized. The signal-to-noise ratio (SNR) of a particular tissue—indicating how much meaningful signal is present compared to background interference—is calculated by dividing the average tissue signal by the standard deviation of the background noise. The sound of the machine and the voices of the radiologist are the first things the person inside the scanner perceives. The narrative MRI produces is an acoustic one. It is not simply auditory though the ears, because listening happens "through the body as a whole and through the technological apparatus (namely, headphones and the bed that vibrates according to the beating rhythm of MRI runs)" (Casini, 2011, p. 85). Reversing Gallagher's understanding of fMRI, the stillness of the person in the scan and the lack of proper visuality of neurological mechanisms are conditions that allow us to dismantle our pictorial gaze in our understanding of brain images as personal representations, once the medial mechanism is taken into account.

It is true that such an enterprise has already been undertaken by posthumanistic thinking (Haraway, 1988), as well as by more recent scholarship in medical visual studies, which argues that brain images should no longer be understood primarily as representations or

signs (Pinotti & Somaini, 2016) but as tools that actively perform functions (Ihde 1998) or as "operative images"—that is, images that do not depict objects but participate in operational processes (Farocki, 2004; Hoel & Lindseth, 2016; Hoel, 2020; Parikka, 2023). However, the epistemic premise implied thereby still relies on the spectator's "looking at". What I am proposing instead is for the whole brain/body/machine embedding to be conceived of as an epistemic-material-medial environment, *made visible* by brain imagers. Brain scans are then far from being transparent. They record biochemical, bioelectrical, physical, metabolic, and magnetic processes occurring at a given moment in some nonvisual matter which is not exposed to light and varying according to the conditions and stimuli to which human beings are exposed, by way of *visibilization* devices (Cappelletto, 2022a, 2022b; Cappelletto & Galimberti, 2023). As Nicole Miglio and Giulio Galimberti have also pointed out, such devices make visible precisely those phenomena "that have nothing to do with the visual realm, nor are they detected by lenses or prostheses that intercept signals from the visible electromagnetic spectrum" (Miglio & Galimberti 2022, p. 71). Accordingly, "brain imaging is not a perfected visual strategy but an image-making practice, technologically heterogenous to previous visual media based on a representational principle" (Cappelletto, 2022a, p. 108). Visibilization makes apparent why there is no transparency as such, since there is no direct access to matter, no point of contact between the image and what is displayed. It follows not simply that nonvisual matter is made visible to the eye, but that a new discourse on individuality opens up. MRI and computer-generated visual reconfigurations of physical data in general enact the brain/body rather than represent it. The very notion of personal identity they configure is therefore entangled, exactly at the moment in which it is pictured. Every form of mediation is itself part of ourselves. Finally, it is not just human beings who behave in medial and cultural environments; machines are too.

REFERENCES

Ahmed, S. (2006). *Queer Phenomenology: Orientations, Objects, Others*. Durham: Duke University Press;

American Psychological Association (2007). *Functional Magnetic Resonance Imaging: A New Research Tool*. Washington, DC: American Psychological Association. Retrieved from https://www.yumpu.com/en/document/read/10408039/functional-magnetic-resonance-imaging-a-new-research-tool;

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics And the Entanglement of Matter And Meaning*. Durham: Duke University Press;

Baudrillard, J. (1986). *America*. Trans. Ch. Turner. London: Verso, 1988;

Bolter, J. D. & Grusin, R. (1999). *Remediation: Understanding New Media*. Cambridge, MA: The MIT Press;

Braidotti, R. (2002). *Metamorphoses: Towards a Materialist Theory of Becoming*. Cambridge: Polity Press;

Cappelletto, C. (2022a). *Embodying Art: How We See, Think, Feel, and Create*. New York, NY: Columbia University Press;

Cappelletto, C. (2022b). Brain Image Making and Cyborg Visuality. *Reti, saperi, linguaggi*, *1*, 87-104;

Cappelletto, C. (2023). On Visibilization. Sight, Living Matter, and the Self. *Reti, saperi, linguaggi*, *2*, 273-290;

Cappelletto, C. (2025). Giochi di identità. Volti, maschere, corpi. In M. Calbi & C. Cappelletto, *Io, lei, noi. Storie di mascherine, facce, persone*. Milano: Mimesis;

Cappelletto, C. & Galimberti, G. (Eds.) (2023). Visually Materializing Nonvisual Matter. Challenges, Paradoxes, and Approaches. *Reti, saperi, linguaggi*, *2*, special issue;

Casini, S. (2011). Magnetic Resonance Imaging (MRI) as Mirror and Portrait: MRI Configurations between Science and the Arts. *Configurations*, *19*, 73-99;
Churchland, P. (1984). Eliminative Materialism. In Id., *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*. Cambridge, MA: The MIT Press;
Churchland, P. S. (2013). *Touching a Nerve: The Self as Brain*. New York, NY: Norton;
Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Putnam;
Deleuze, G. (2024). *Sur Spinoza: Cours, novembre 1980-mars 1981*. Paris: Les Editions de Minuits;
Deleuze, G. & Guattari, F. (1980). *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis-London: University of Minnesota Press, 1987;
Dumit, J. (2003). Is It Me or My Brain? Depression and Neuroscientific Facts. *Journal of Medical Humanities*, *24*(1-2), 35-47;
Ehrenberg, A. (2008). Le Cerveau 'social': Chimère épistémologique et vérité sociologique. *Esprit*, *1*, 79-103;
Farocki, H. (2004). Phantom Images. *Public*, *29*, 12-24;
Ferret, S. (1993). *Le philosophe et son scalpel. Le problème de l'identité personnelle*. Paris: Minuit;
Foucault, M. (1963). *The Birth of the Clinic: An Archaeology of Medical Perception*. London: Tavistock, 1973;
Galimberti, G. & Miglio, N. (2022). Intra-Active Sense-Making Towards a Performative Understanding of Biomedical Imaging. *Cinéma & Cie*, *22*(39), 69-82;
Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience*, 7, 443. DOI: 10.3389/fnhum.2013.00443;
Gallagher, S. (2020). *Action and Interaction*. Oxford: Oxford University Press;
Gallagher, S., Hutto, D., Slaby, J., & Cole, J. (2013). The brain as part of an enactive system (commentary). *Behavioral and Brain Sciences*, *36*(4), 421-422;
Gallese, V. & Morelli, U. (2024). *Cosa significa essere umani? Corpo, cervello e relazione per vivere nel presente*. Milano: Cortina;
Giannachi, G. (2023). *Technologies of the Self-Portrait. Identity, Presence and the Construction of the Subject(s) in Twentieth and Twenty-First Century Art*. New York, NY: Routledge;
Hixson, J. R. (1983). New Seeing-Eye Machines... Look Inside Your Body, Can Save Your Life. *Vogue*, July, 238-239, 254-256;
Hoel, A. S. A. (2020). Images as Active Powers for Reality. A Simondonian Approach to Medical Imaging. In E. Alloa & C. Cappelletto (Eds.), *Dynamis of the Image. Moving Images in a Global World* (pp. 287-310). Berlin & Boston: De Gruyter;
Hoel, A. S. & Lindseth, F. (2016). *Differe*ntial Interventions: Images as Operative Tools. In K. Kuc & J. Zylinska (Eds.), *Photomediations: A Reader* (pp. 177-183). London: Open Humanities Press;
Horikawa, T. & Kamitani, Y. (2017). Generic Decoding of Seen and Imagined Objects Using Hierarchical Visual Features. *Nature Communications*, *8*. doi.org/10.1038/ncomms15037;
Ihde, D. (1998). *Expanding Hermeneutics: Visualism in Science*. Evanston, ILL: Northwestern University Press;
Latour, B. (2004). How to Talk About the Body? The Normative Dimension of Science Studies. *Body and Society*, *10*(2/3), 205-229;
Lauterbur, P. C. (1973). Image Formation by Induced Local Interactions Examples Employing Nuclear Magnetic Resonance. *Clinical Orthopedics and Related Research*, 1989, *244*, 3-6;
Parikka, J. (2023). *Operational Images: From the Visual to the Invisual*. Minneapolis, MN: University of Minnesota Press;
Pinotti, A. & Somaini, A. (2016). *Cultura visuale. Immagini, sguardi, media, dispositivi*. Torino: Einaudi;
Ricoeur, P. (1990). *Oneself an Another*. Trans. Chicago, ILL: University of Chicago Press, 1992;
Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson;

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). A second-person neuroscience in interaction. *Behavioral and Brain Sciences 36*(4), 441-462;
Simmel, G. (1901). Die ästhetische Bedeutung des Gesichts, in *Der Lotse. Hamburgische Wochenschrift für deutsche Kultur*, *I*(2), 280-284 [GSG 7, pp. 36-42];
van Dijck, José (2005). *The transparent body. A cultural analysis of medical imaging*. Washington, DC: University of Washington Press;
Vidal, F. (2009). Brainhood, Anthropological Figure of Modernity. *History of the Human Sciences*, *22*(5). DOI: 10.1177/0952695108099133;
Wickens, A. P. (2015). *A History of the Brain: From Stone Age Surgery to Modern Neuroscience*. New York, NY: Psychology Press;
Wittgenstein, L. (1972). *On Certainty*. San Francisco, CA: Harper & Row;
Wittgenstein, L. (1989). Philosophie. Philosophy. In J. Klagge & A. Nordmann (Eds.), *Philosophical Occasions 1912-1951*. Indianapolis, IND & Cambridge: Hackett 1993.

LISA SANGUINETI
Università degli studi di Roma Tor Vergata
Vita-Salute San Raffaele University
lisa.sanguineti.sv@gmail.com

RECLAIMING THE SELF. REDEFINING AND AFFIRMING MARGINALIZED IDENTITIES THROUGH CONTEMPORARY ART

abstract

Throughout history, artists have used their work to investigate and express their identity (whether individual, community, political, sexual or social). Artists belonging to historically marginalized communities have used their artistic practices as a weapon in response to stereotyping and oppression. The present paper intends to present different contemporary artistic experiences focused on the theme of marginalized identities and their investigation within a decolonial approach. An attempt will be made to investigate how contemporary artists are positioning themselves in processes of identity inquiry, claiming, and reappropriation. A comparison will be made between different artists that have decided to use their art as a tool for reclaiming their own identity, cleansing it of any remnants of stereotype previously produced by dominant cultures. The 60th Venice Biennale will serve as a main case study to examine how contemporary artists are dealing with the notion of identity.

keywords

contemporary art, identity, decolonization, Venice Biennale

Phenomenology and Mind, n. 29 - 2025, pp. 152-160
DOI: 10.17454/pam-2913
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

Throughout the eras, artists have investigated through their artistic practices the issues of identity, whether it is conceived as individual, community, political, or social. Through their work, artists have had the chance to cathartically analyze the self and express it through different creative forms. Sometimes the creative product represents the result of the investigation, while at other times it serves as an integral part of the investigative process itself. In response to the systemic oppression, homogenization, and censorship historically perpetrated by the dominant culture, artists and women artists from marginalized communities claim and celebrate their own identities while spreading awareness on established stereotypes and contributing to their demolition. Therefore, decolonial art practices have become for artists engaged in social and cultural decolonization a tool for identity redefinition and self-determination. Through their art, contemporary artists who belong to socially and politically marginalized communities have finally had the chance to express and affirm their own identity, without having it mediated by the Western gaze[1]. The present paper intends to present different contemporary artistic experiences focused on the theme of marginalized identities and their investigation within a decolonial approach. An attempt will be made to investigate how contemporary artists are positioning themselves in processes of identity inquiry, claiming, and reappropriation.

In the first paragraph of the paper the role of contemporary artists in these processes will be analyzed. A comparison will be made between different artists that have decided to use their art as a tool for reclaiming their own identity, cleansing it of any remnants of stereotype previously produced by dominant cultures. Through the short analysis of the art of Robert Colescott, an unusual way of fighting stereotypes through painting will be presented. Secondly, the practice of portraiture will be investigated through the art of Lynett Yiadom Boyake.

The second section of the paper will address the 60th Venice Biennale as the main case study. As decolonization and inclusion were core themes of the *Foreigners Everywhere*

1 In this process of abandoning the Eurocentric approach for the qualitative evaluation of art and establishing new perspectives other than the Western one, artivism plays a key role. Art practices related to individual artist-activists or creative collectives have facilitated processes of cultural decolonization and re-signification of heritage. On the role of artivism in the decolonization processes see, for instance, Castellano (2022).

exhibition, this section will present and analyze the artists who have explored their identities and presented their artworks within such an historically elitist setting.

2. Redefying identity through contemporary art

2.1 A premises: Sonya Boyce and the risk of generalizing

Before delving into the study of how contemporary artistic practices have addressed the analysis of marginalized identity, a preliminary remark is necessary. In the wake of postcolonial studies and the popularity of the inclusive discourse, one could fall into error by making simplifications and generalizations. Just because an artist belongs to a marginalized community, it doesn't mean that they are always willing to represent their people and their condition through his art. This issue is well articulated by the British Afro-Caribbean artist Sonia Boyce, who has represented Britain at the 59th Venice Biennale. When asked about her identity, Boyce stated:

I am a black woman artist. That's not necessarily who I am, but what I am. And so I always feel nervous when people expect that what one is represents what one does... and represents anyone else who could come under that rubric (Diawara, 1992, p. 192).

Boyce herself has underlined that the presence of a few Black artists in the country does not imply that they wish to act as spokespeople for their entire community (Diawara, 1992, p. 193)[2]. Speaking about herself, Boyce has claimed that she's trying distance herseld from the notion of identity, as it's a limiting concept that can trap you:

If I speak, I speak "as a" black woman artist or "as a" black woman or "as a" black person. I always have to name who I am: I'm constantly being put in that position, required to talk in that place ... never allowed to speak because I speak. I want to find out what other things I can talk about. I no longer want to describe who I am (Diawara, 1992, p. 195).

Boyce's perspective allows us to understand that, when discussing identity, it is essential to consider the complexity and the individuality of each artist- Her reflection highlights the risks of oversimplification that stem from assumptions and generalization.

2.2 Robert Colescott and the use of stereotypes

Since the second half of the XIX century, artists have increasingly challenged and dismantled the legacy of colonialism on cultural, social and personal identity. For years, marginalized communities, such as black or indigenous people, have been represented through stereotypical images, mediated by a dominant culture. The products of these stereotyping were artificial and false identities, images produced from the outside, shaped by prejudice and presumption. Even though these images were belittling and did not correspond to reality, they became established in pop culture and have not been eradicated yet.

The black community is among those witnessing the worst stereotyping from the dominant (white) culture. Sterotypical representations were used since the colonial period and towards the Jim Crow-era to depict black people, emphasizing physical features and making them displeasing and laughable. Figures such as Aunt Jemima, Jim Crow or Sambo are perfect examples of this made-up black identity, built by white people to slander the black community[3]. Their aim was propagandistic and contributed to furthering the

2 "There have been quite a few black artists producing work. Now, because of the pressures within the fine arts world, and the additional challenge of being a black male or female artist working in the system, it's very very difficult to survive and to develop as an artist. And when attention has been paid to black British art, there hasn't really been a discussion of things *around* the work. But I am not the only person; I am just one of the people who are quite well known" (Diawara, 1992, p. 193).

3 For an in-depth study on the stereotyped image of black women, see Sewell (2013).

separation between communities[4]. Thanks to postcolonial studies and the raising awareness on cultural affirmation and fair representation, contemporary artists from marginalized communities have finally had the chance to claim their own space. Through their artworks, they have started a proper process of decolonization of the arts and historical narratives, by reinterpreting historical events, reclaiming ownership on cultural heritage and promoting inclusivity and diversity[5].

Robert Colescott attempted to use the "artificial identities" of the sadly well-known black stereotypical characters to claim a new space for blacks in contemporary art history. Colescott was the first African American painter to represent the U.S. at the Venice Biennale, in 1997, and the first black artist to have his work published on the cover of Artforum in 1984 (Gallant, 2022). His artworks consist of rearrangements of famous western paintings, yet they are filled with only black stereotypical figures[6]. His painting is based on a humorous complaint that turns out irreverent and unsettling for the public (especially the white one).The investigation on black identity intrigues and captures the viewer, thanks also to the cartoonlike figures and the bright colors used by the artist. However, he creates in the audience a mix of feelings: comic features that first seem entertaining suddenly leave room for a sense of guilt and discomfort. The artist himself used to say: "If you decide to laugh, don't forget the 'humor is the bait,' and once you've bitten, you may have to do some serious chewing. The tears may come later" (Copeland, 2009). In one of his masterpieces, *George Washington Carver Crossing the Delaware: Page from an American History Textbook* (1975), Colescott exhibits this artificial black identity created by whites. The Emanuel Leutze's work *Washington Crossing the Delaware* is reproduced on Colescott's canvas, but here the main protagonists are replaced with Sambo, Aunt Jemima, Jim Crow and other stereotyped black figures. Through this artwork, Colescott rewrites United States history as written by blacks, and he reclaims a space for black people in the contemporary art world. Starting from his identity as a black male, he reshapes the Eurocentric narratives of art history, while also advocating for fair representation of marginalized cultures. In his artworks, Colescott often investigates and plays with both black and white identities. *Shirley Temple Black and Bill Robinson White* is another example of the "swap game" played by the artist: in the painting he switches the racial identities of the two well-known actors to create amusement and bewilderment in the audience. The questions that Colescott asks the audience are clear: can identity influence the way a person is perceived and loved by the public? Would a black Shirley Temple have been loved as much as the white one?

2.3 Portraying blackness

Portraits and self-portraits have been major tools for the investigation and representation of personal identities. Throughout history artists have used these forms of art to investigate themselves and capture their true selves, or to represent all the possible facets of other human beings. However, contemporary artists have found different ways of employing portraits and self-portraits to provide a decolonized perspective on identities to the public. Kehinde Wiley, for example, is known worldwide for his baroque portraits that sumptuously celebrate blackness and the urban culture often associated with it. But is blackness an identitarian feature for every contemporary portraitist?

4 On the propagandistic aim of stereotyping black communities, see Melson–Silimon, Spivey, & Skinner–Dorkenoo (2023).

5 The so called "Fanonian art Practices", for example, built on the decolonial perspective of the psychiatrist and philosopher Frantz Fanon and consist in artistic research that start from the remnants of blackness and that seek to escae the constraints of the western culture through inclusive and decolonial creative practices. See Nkosinkulu, 2024.

6 On the life and work of Robert Colescott throughout his career, see Platow, Stokes Sims, & Weseley (2019).

Those painted by the British Ghanaian artist Lynette Yiadom-Boakye, for example, are unusual, imaginary portraits. In her artworks Yiadom-Boakye represents daily life scenes or portraits of people who do not exist and that she can only see in her mind. The figures that fill her works are black people and their outlines are blurred like those of a distant memory. This lack of clarity, the use of generic clothes and unidentifiable backgrounds allow her to portrait figures that are timeless and spaceless. The color of her artworks, generally mute, and the vague titles (such as *To Improvise a Mountain* or *Solitaire*) contribute to this imaginative aura. The representation of black people and the definition of black identity is a matter of normality for the artist: since she was raised by black people as a black person, she can only paint black people (Higgie, 2012). Yiadom-Boakye stated "Blackness has never been other to me. Therefore, I've never felt the need to explain its presence in the work anymore than I've felt the need to explain my presence in the world, however often I'm asked"[7]. The difference between Yiadom Boakye and Colescott is evident: while the American painter claims a space for black people in art history by using sarcasm and a captivating figurative language, Yiadom Boakye's research is much more intimate, subtle, and personal. Like Boyce, Yiadom Boakye complains about expectations placed on Black artists to explain or justify their work in relation to their race. The decolonial approach to Black identity is rooted in the artist's decision not to use whiteness as a paradigm: she describes blackness and her identity as they are, without defying them by negation (as non-white). The blurred contours and the fictional characters of her paintings allow her to explore not only blackness, but its "infinite possibilities" (Richler, 2023).

3. The 2024 Venice Biennale: between decolonizing and reclaiming identities

Adriano Pedrosa, the curator of the 60th Venice Biennale decided to make the concept of identity one of the core themes of the renowned exhibition. With the title *Stranieri ovunque. Foreigners everywhere*, inspired by a famous series of work by the collective Claire Fontaine[8], the exhibition aimed to mirror the complex social and political context of the present. Artists were asked to reflect on the global challenges of our society, in which people can no longer be classified through borders, languages, genders or sexuality. According to Pedrosa, the condition of the foreigners, the strangers, starts within the person, who "will always feel foreigner", no matter where or with whom: that of the foreigners is an inherent, inevitable condition that, at the same time, unites all humanity (Pedrosa, 2024b, p. 54). The effort to decolonize the Venice Biennale and to explore the concept of identity in a broader and more inclusive way was evident. By doing so, the curator aimed to allow the artists and their communities (that mostly join the Biennale for the first time) to speak and advocate for themselves. Thanks to this approach, the exhibition is enriched with a multitude of approaches and perspectives that finally make it possible to anyone in the audience to feel represented, while fostering the dialogue between people and cultures.

The exhibition hosted artists who, for a variety of reasons, were connected to the condition and experience of foreignness. Queers, indigenous, naive, immigrant artists have pondered the challenge proposed by Pedrosa by exploring the themes of migration, displacement and identity. In the following section of the paper, different ways of

7 The quote is attributed to the artist by the Tate Britain, and it has been used for the presentation of the artworks. Retrieved from: https://www.tate.org.uk/art/artists/lynette-yiadom-boakye-16784/introduction-lynette-yiadom-boakye. For other personal reflection by the artist on blackness and its meaning in her artworks, see Yiadom-Boakye, 2014.

8 The artworks consist in neon lights in different colors displaying the phrase "Foreigners everywhere". The series was initiated by the collective in 2004 and addresses the issues of identity and migration within the contemporary socio-political context. On Claire Fontaine and its theoretical techniques and procedures see Chiari, 2024.

addressing the issues of identity, belonging and foreignness within the 60th Biennale will be analyzed[9].

3.1 Decolonized national identities

Pablo Delano's project, which spread across an entire hall of the central pavilion at the Giardini, broke down, through a decolonial approach, the concept of community, individual, political and social identity. The artist's reflection started from the complexities of his home country, Puerto Rico, which since the arrival of Cristoforo Colombo has been living under colonial oppression. After the Spanish-American war, in 1898, the island became an unincorporated state and has still been striving to gain its independence. In his project, that has been already presented on other occasions, Delano decides to build a sort of museum within the museum and unfolds a multilayered and harsh systemic critic. The work derives its title, *The Museum of the Old Colony*[10], from a popular American soft drink popular in Puerto Rico from the 50s and still on the market. The title alludes with bitter irony to the island's past and present political conditions In his installation/project Delano exhibits historical photographs of Puerto Rico, mostly taken by U.S. photographers and depicting a variety of subjects, events and scenes. The photos that are shown are divided in different sections, depending on the themes, and are combined with videos and everyday objects from the island. Through all these elements, Delano traces a history of the cultural and political colonization witnessed by Puerto Rico. He reflects on the concept of museums and the extractivist approach of those who have stereotyped, colonized and exploited the Caribbean communities. Viewers question themselves on their potential role in this perpetual act of colonization. The project challenges the concept of identity: after a history and a present of colonization, can there truly be an authentic Puerto Rican identity? Furthermore, Delano questions the main mechanisms and rules of the Biennale, an exhibition that has been historically based on national pavilions: since Puerto Rico doesn't hold the state of a nation, it could never be represented by an artist in a national pavilion. Despite this "formal" contradictions, Puerto Ricans have a sense of community and nationality that unites them, a common Puerto Rican identity that is based on the same language, the same cultural references, symbols and practices. The identity that Delano is trying to exhibit and bring to a bigger public is an identity that is not politically recognized yet, but still exists and vibes in the present days[11].

3.2 Mute and muted identities

The decolonized approach towards the investigation of identity can also be seen in another specific artistic production dedicated to the analysis and reflection on gender and sexuality. Usually, the aim of these practices is to decolonize the gaze, to affirm other ways of living, of being, of experiencing sexuality and gender identity that differ from those normally accepted. For the 60th Venice Biennale Gabrielle Goliath presented an immersive installation

9 The descriptions of the artworks in the following paragraphs are based not only on the study of the cited sources but also on the firsthand experience of the author during their visit to the exhibition in April and May 2024.

10 On Pablo Delano's *Museum of Old Colony* see Berger (2020) and Katzman (2022).

11 Palo Delano was not the only one proposing a reflection on decolonized political and national identities at the 60th Venice Biennale. Among the artists presented at the exhibition, for example, The Unknown Chilean artist (Arpilleristas) proposed, through their tapestries, a decolonized perspective on Chile (Pedrosa, 2024b, p. 81). In the Danish pavilion, Inuute Storch traced the colonial history of Greenland from a colonized perspective (Pedrosa, 2024b, p. 389), while in the spaces of the USA Pavilion Jeffrey Gibson proudly celebrated indigenous identity against oppression (Pedrosa, 2024b, p. 449).

dedicated to giving space to marginalized identities. For her project, *Personal Accounts*[12], which started in 2014 and has been brought all over the world, the Sud African artist decided to interview black people, people of color, queers, transgenders and indigenous. The interviewed express all the limits of the political and social systems that oppress them: patriarchal logics, binarism, colonial attitudes, racism. In a choral catarchical act, people of different cities and countries of the world have shared their stories with the artist, telling her about their traumatic experiences. They told their stories to the artist, united under the same condition as survivors of an oppressive, discriminatory, and bigoted system. However, the artist has decided, with the prior consent of all the participants, to cut and delete all the spoken words: only sighs, whispers, murmurs, waverings, swallowings, and pauses echo in the room where the videos of the interviews are screened. As a result of these choices, viewers find themselves totally immersed in a profound chanting of hesitation and surrounded by the presence of the participants and their discomfort. By forgoing dialogues, the artist creates a universally understandable language through which the participants express their feelings without any fear of being misunderstood and not trusted. That of Goliath is a sensitive and delicate way to investigate marginalized identities and to reclaim a space for their conditions in the public dialogue. The lack of words acts as a metaphor of the lack of comprehension that these people have experienced in their lives. The viewers question themselves on their past attitudes towards marginalized communities and challenge the concepts of identity (whether individual, social, or sexual) that they have normalized throughout their lives.

3.3 Reclaiming identity through the group: The Aravani Art Project

A common and widespread mistake is to conceive identities as solely individual. Many of the projects presented at the 60th Venice Biennale have shown that many artists can relate the sense of belonging and the concept of identity to their community: the individual defines himself through the group and, at the same time, is empowered by the collectivity. The importance given to plural projects and artistic collectives was crucial for the exhibition, as can easily be seen from two elements: the title of the exhibition itself, retrieved from the artwork of a collective, Claire Fontaine, and the mural on the façade of the central pavilion, realized by the Brazilian *Movimento Dos Artistas Huni Kuin*.

Sometimes working as a group empowers marginalized individuals, leading them to speak up for themselves, as in the case of the Aravani Art Project. The Indian collective is made of cis and transgender women and was created with the specific purpose of spreading positivity and hope among their community (Pedrosa, 2024b, p.79). The bright and colorful murals contribute to the process of affirmation and acceptance of the trans community within India. By showing in their mural trans people in a gigantic scale, Aravani Art Projects wants to metaphorically overcome the traditional concept of identity and put an end to stereotyping and stigmas. The collective acts as a safe space in which everyone can express its true self and unveil their true self[13].

4. Conclusion

Despite the attempt of numerous communities and individuals to create a more inclusive society, episodes of racism, oppression and ostracism are still sadly present in our society.

12 On the project's website, *Personal Accounts* is described as a "transnational, decolonial, black feminist project of repair". See https://www.personal-accounts.art/personal-accounts.

13 The Aravani Art Project's mission, vision and methodology can be further explored on the collective's website, see https://aravaniartproject.com/.

For this reason, it is important that the artistic research presented in this paper gains more and more visibility and importance in public discourse. That of Pedrosa is one of the potential attempts to achieve this goal: the 60th Venice Biennale exemplifies a progressive shift towards a broader notion of identity. Thanks to this prestigious platform, artists were not only able to present their work for the first time at one of the world's most important industry events but, more importantly, to share new narratives and perspectives that move away from traditional Eurocentrism. As we look to the future, these artistic experiences and their new-found popularity offer a beacon of hope. In any case, as Boyce well expressed, each artistic practice will have to be valued in its specifics and contextualized. By embracing and celebrating the diverse range of human experiences and fostering a broader concept of identity, society will move towards greater inclusivity. In this collective effort, art will play a fundamental role, serving as a catalyst for change and an amplifier of voices.

REFERENCES

Berger, M. (2020). *The Museum of The Old Colony*. Center for Art, Design and Visual Culture; Baltimore, MD: University of Maryland;

Castellano, C. G. (2022). *Art Activism for an Anticolonial Future*, New York, NY: State University of New York Press;

Chiari, A. (2024). *A User's Manual to Claire Fontaine*. Milan: Lenz Press;

Copeland, H. (2009). Truth to Power. Huey Copeland on Robert Colescott (1925-2009). *Artforum*, 48(2), October, 59-60;

Diawara, M. (1992). The Art of Identity. *Transition*, 55, 192-201. DOI: 10.2307/2934865;

Gallant, L. (2022, April 1). How Robert Colescott used art history to force viewers to confront their prejudices. *The Art Newspaper*. Retrieved from https://www.theartnewspaper.com/2022/04/01/robert-colescott-art-and-race-matters-retrospective-chicago;

Higgie, J. (2012, April). A life in a day. The fictitious portraits of Lynette Yiadom-Boakye. *Frieze Magazine*, April 2012, Issue 146. Retrieved from https://www.frieze.com/article/life-day;

Katzman, L. (2022). *The Museum of the Old Colony. An Art Installation by Pablo Delano.* Harrisonburg, VA: James Madison University;

Melson-Silimon, A., Spivey, B. N., & Skinner–Dorkenoo, A. L. (2024). The construction of racial stereotypes and how they serve as racial propaganda. *Social and Personality Psychology Compass*, 18(1), e12862. DOI: 10.1111/spc3.12862;

Moss, D. (2020). "Radical Empathy": Amy Sherald's Portrait of Michelle Obama. In T. Caragol, D. Moss, R. J. Powell, & K. Sajet (Eds.), *The Obama Portraits*. Princeton, NJ: Princeton University Press;

Nkosinkulu, Z. (2023). Fanonian Art Practices: Towars A Decolonial Grammar of Being. *Journal of Black Studies*, 54(5), 394-409. DOI: 10.1177/00219347231173151;

Pedrosa, A. (2024a). *BIENNALE ARTE 2024. Stranieri Ovunque. Catalogo della 60. Esposizione Internazionale d'Arte.* Venice*:* La Biennale di Venezia;

Pedrosa, A. (2024b). *BIENNALE ARTE 2024. Stranieri Ovunque. Guida Breve della 60. Esposizione Internazionale d'Arte*. Venice: La Biennale di Venezia;

Platow, R., Stokes Sims, L., & Weseley, M. (2019). *Art and Race Matters: The Career of Robert Colescott*. New York, NY: Rizzoli Electa;

Richler, P. (2023, March 4), Many Infinities: The Vital Canvases of Lynette Yiadom-Boakye. *The Quietus*. Retrieved from https://thequietus.com/culture/art/lynette-yiadom-boakye-fly-in-league-with-the-night-tate-britain-review-2/;

Sewell, C. J. P. (2013). Mammies and Matriarchs: Tracing Images of the Black Female in Popular Culture 1950s to Present. *Journal of African American Studies*, 17(3), 375-392;

Ulrich Obrist, H. (2012). Hans Ulrich Obrist Interviews the Afropolitan Lynette Yiadom-Boakye. *Kaleidoscope*, Summer 2012, Issue 15, 98-103;
Yiadom-Boakye, L. (2014). Artist statement. *Callaloo*, 37(4), 903-906.

Links
https://aravaniartproject.com/;
https://www.personal-accounts.art/personal-accounts.

DILETTA CAIMMI
Vita-Salute San Raffaele University, Milan
diletta.caimmi@gmail.com

PLURAL IDENTITIES AND COLLECTIVE AUTHORSHIP: GROUP EXPERIENCES IN CONTEMPORARY ART

abstract

This paper intends to provide some historical and philosophical insights to reflect upon the enduring impact of group experiences which, since the post-war period, have constituted an important peculiarity in the landscape of contemporary arts. We will reconstruct some of the reasons that have led artists to seek new paths of subjectivation through collective creation: moving across the technological experimentation of the groups of the 1960s, the political militancy of the collectives of the 1970s, the format of the artistic acronym of the 1980s and the new possibilities opened up by the Net in the 1990s, we will retrace the deconstruction of the romantic mythology of the author, as it has traditionally been consolidated. By emphasizing shared creative processes, these practices propose innovative models of identity and subjectivity that transcend individual boundaries and reflect the complex interplay of technology, politics, and social interaction in contemporary art.

keywords

plural artist, artistic collectives, contemporary art, identity, authorship

Phenomenology and Mind, n. 29 - 2025, pp. 162-172
DOI: 10.17454/pam-2914
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction: The Plural Artist

This contribution aims to propose some insights, without any claim to exhaustiveness, on group experiences that, through collaborative and horizontal artistic practices, have dismantled the traditional notion of authorship. Forms of collective artistic identities have indeed allowed contemporary art to act as a catalyst for new processes of subjectivation, triggering broader debates on the conventions that have historically shaped our conception of identity. The framework informing this article draws on a theoretical paradigm borrowed from the humanities and social sciences, which, over the 20th century, disavowed the idea of a presumed naturalness of identity. While the traditional "ontological perspective" conceives identity as fixed, metaphysically guaranteed, and independent of human action, the "sociological perspective" considers identity as fluid, shaped by social recognition, negotiation, and competition: in the first case, identity can just be discovered and contemplated, in the second, it is instead continually invented, constructed, and redefined in the social arena (Remotti, 2010).

Not only several collaborative experiments have recognized the underlying material and authorial co-dependence in an artistic project, but there are also significant groups that have made this recognition an explicit and meaningful poetic hallmark. We will use the notion of "plural artist" (Balzola & Rosa, 2011) to talk about a constellation of possibilities connected by the choice to create a single authorial device, a unique "artistic identity" starting from the coexistence of many individualities; we will use some instances put forward by what Argan termed the "reasons of the group" (Argan, 1963) to question how a certain articulation of artistic practice can provide practical and imaginative tools for conceiving a plural identity. Furthermore, we will examine how collective creation can delineate new models of subjectivity that challenge the consumer-oriented subjectivity promoted by the neoliberal order, which tends towards a privatized conception of material and immaterial resources.

The desire to explore these issues stems from an observation: in the last decade, there has been a growing interest in artistic collectives, increasingly enhanced by innovative curatorial proposals and highlighted within important institutional spaces. Plural artists have recently been entrusted with the curatorship of some of the most prestigious international exhibitions in Europe, such as the 2016 Berlin Biennale, curated by the New York collective DIS[1], or

1 The group is characterized by an interdisciplinary approach to advertising, fashion, communication, and popular culture via the Internet: born in 2010 from a virtual and interactive magazine, it has over time proposed cross-media

Documenta Fifteen, curated in 2022 by the Indonesian collective ruangrupa[2]. Plural artists have been awarded prestigious honors such as the Turner Prize, organized by the Tate Gallery and reserved for young British artists: in 2021, the shortlist of finalists consisted exclusively of collectives – Array Collective, Black Obsidian Sound System, Cooking Sections, Gentle/Radical, and Project Art Works; the award was given to Array Collective, which works to create collective actions in response to socio-political issues impacting Northern Ireland. In 2024, on the occasion of the 60th International Art Exhibition in Venice, Mataaho Collective won the Golden Lion as best participant; the title of the Biennale itself, *Foreigners Everywhere*, was inspired by a series of works produced by a plural artist, composed of Fulvia Carnevale and James Thornhill, which speaks of herself in the singular and identifies as Claire Fontaine: born in 2004 in Paris, two years later she releases an "announcement" to the public of the French television broadcast *Ce soir (ou jamais!)*. In presenting herself, she writes:

> Claire Fontaine is a groupuscule made up of groupuscules, because each one of us is already several in themselves. We work under a pseudonym because our bodies are not the bodies of the authors of this work, they are the receptacles of collective ideas and of the political problems that traverse us, but we are not just couriers. We collaborate and participate. We don't direct anything [...]. We are like you: whatever singularities caught in the cogs of global capitalist misery. The only thing that differentiates us is the fact that the general deafness haunts us day and night like a buzzing in our ears at a painful frequency. (Fontaine, 2020, pp. 75-76)

The growing visibility that artistic group projects enjoy today suggests that the legacy of past collective experiences has been recognized and assimilated, and that the individual's choice to found themselves in the plural subject has delineated new possible identities worth analysing.

2. 1960s and 1970s: Plural Identities Between Technological Experimentation and Political Militancy

With the historical avant-gardes of the first decades of the twentieth century, an enhancement of group work has emerged within the production and contemporary art criticism, reaching its peak in the years following the Second World War. The development of consumer society set the stage for a proliferation of artistic groups, showing how artistic practice was perceived as a potential field of resistance against the alienating logic of massification and cultural homogenization. On this point, the reflection proposed by Argan (1963) in the above-mentioned article is paradigmatic. He states: "Whoever wants to defend the free activity of the individual from the torpid and lethal inertia of the mass must reflect, first of all, that the

and cross-platform interventions both online and offline, supporting new creative practices. DIS's virtual community remains active online, pursuing the goals that have animated the project from the beginning: "We started dis.art with the idea that complex thinking about politically challenging issues might be better served through entertainment than academic papers [...]. Rather than compete for clicks and ad revenue on YouTube, we distribute directly to individuals as well as libraries, universities, and art institutions. We aim to foster a solution-driven community, and like you, our desire is to rethink the way social, political, and economic structures operate, and to imagine generative blueprints outside of dominant narratives. The future demands not just our expertly honed skills in watching, but also imagining and understanding" in https://dis.art/ (accessed on 05/06/2024).

2 Although the artistic direction of Documenta was entrusted to a core group of ten members, the total number of actual members of ruangrupa is not fixed. The official communication related to the exhibition was handled by the street newspaper Asphalt, sold by homeless people in the cities of Kassel, Göttingen, and Hanover. The first statements disseminated in the press materials outlined the centrality of the concept of *lumbung* ("communal rice barn"), which for the entire duration of the exhibition served as a theoretical and operational paradigm for creating relationships among all the involved realities. For more information see https://documenta-fifteen.de/en/lumbung/ (accessed on 05/06/2024). Some of the most interesting publications about the ruangrupa curatorial practice are included in Kolb & Richter, 2022.

fundamental quality of the human person is the capacity, the will, to put oneself in common, to associate with others for a common purpose, to coordinate one's actions with others, to form a group" (Argan, 1963).

During the 1960s, as aesthetic sensibility gradually shifted its focus from fixed forms to movement (the work is no longer *a thing*, but *a field of events*), from passive contemplation to active spectatorship, from static enjoyment to a broader phenomenology of perception, a constellation of collective projects spreads worldwide. From the second half of the 1950s to the first half of the 1960s in Italy we witness the birth of Gruppo T and MID in Milan, Gruppo N in Padua, Gruppo Uno in Rome, Gruppo ATOMA in Livorno, Gruppo 63 in Palermo, Gruppo Tempo 3 in Genoa, and Gruppo V in Rimini; in France, there is the Groupe de Recherche d'Art Visuelle and Equipo 57, formed by Spanish artists in exile during the Franco dictatorship; in Germany, Gruppo Zero; in Munich, Gruppo 47; in the Netherlands, Gruppo Nul; in Moscow, Dvijzenije; in Cleveland, the Anonima Group; and in Osaka, the Gutai Group[3].

The desire to form a group responds, on one hand, to a certain dissatisfaction with an art market traditionally based on great personalities embodying a certain "style", and on the other, to the need to rethink collectively the position of artistic practice within the technological horizon emerging from the economic boom, and within the pervasiveness of technique that defines new lifestyles[4]. It is no coincidence that many collectives born in this period are characterized by a strong techno-experimental inclination and interest in what would be defined as kinetic art or arte programmata: the aesthetic object is based on a computational program that allows for formal, chromatic, and perceptual variation of the space and the visual sequences therein, playing on repetitions, variations, and random events.

From this perspective, since 1961, the history of several groups intertwines around the periodic exhibitions of Nove Tendencije (New Tendencies), an international organization managed by some artists based in Zagreb, around which the avant-gardes related to kinetic art gather. Returning to the memory of that experience on the occasion of an exhibition in 2013, the artist Getulio Alviani elaborates a reflection that can serve as an a posteriori manifesto for the theoretical atmosphere in which NT's eight years of activities unfolded:

> The New Tendencies was a very innovative system in everything:
> non-cult of personality (sometimes even anonymity)
> non protagonism
> not commercialization
> not private galleries (but only cultural institutions)
> not elitist art
> not fetishism
> not unique artwork (but only the beginning of multiples for a social purpose)
> not interpretation
> not metaphor
> not mystification
> not strategy
> not... (Alvani, 2013, p. 19)

3 For a more specific overview of some of these groups (with a particular focus on Italian experiences), see Meloni, 2020.

4 In this context, it is again Argan (1961) who lucidly emphasizes, in the introduction to the catalogue of the 12th Lissone Prize in 1961, that "the task of contemporary art is not to preserve at all costs the values improperly called 'humanistic' in a society that becomes increasingly technocratic, but to determine what will be, in the historical figure of modern culture, the place and function of aesthetic activity" (p. 13).

Many people around NT prefer to refer to themselves as "operators" rather than "artists". The typical openness of the 1960s toward the epistemological horizons opened up by new electronic technologies and early forms of automation influences the growing deconstruction of the artist as a single authorial source, considering that the creative use of technologies imposes teamwork among different technical expertise to achieve an artistic result.

This point would be crucial for the new challenges posed three decades later by the integration of virtual reality into aesthetic practice, as attested by the document *For a New Cartography of the Real*, drafted by some Italian scholars in 1993 and conceived as a preliminary draft for the elaboration of a *Manifesto of Art and Communication in the Virtual Age*, among whose founding themes is "Collective Author: new technologies re-launch the possibility and necessity of a collective artistic process". The need to conceive the collective authorship is thus argued:

> The figure of the single artist as the single architect of the artistic process, alpha and omega of the value of the work, the unique repository of the paths of meaning, is transformed. The author is born as a connected, expanded individual, a neuron that receives, activates, translates, interprets, and retransmits information in continuous connection with other entities. It is the driving element of the artistic-communicative process, which is defined in the unceasing interaction between co-authors, disciplines, technologies, and users[5].

The use of technology re-launches the need to broaden the meshes of artistic authorship, conceiving new rhizomatic forms of collaboration within single authorial devices: if it is difficult to transmit, at the level of common imagination, the complex interaction between the various parts of a single artistic organism, it is also due to the absence of appropriate semantic nuances in our language. As Claire Bishop (2012) clearly points out, while "the worlds of music, film, literature, fashion, and theatre have a rich vocabulary to describe co-existing authorial positions (director, author, performer, editor, producer, casting agent, sound engineer, stylist, photographer), all of which are regarded as essential to the creative realization of a given project" (p. 9), the world of visual arts lacks an analogous lexical variety.

With the 1970s, collective configurations took on a militant and activist character, riding the wave of the 1968 protests – partly converged into the Movement of '77. Simplifying, a clear line of distinction can be drawn between the group of the 1960s, a closed formation structured around the programmatic intentions of few artists, operating within the canonical "spaces of art", and the collective of the 1970s, an open platform that embraces the demands of political collectives and acts practices of disorientation, counter-information, and urban reappropriation, favouring actions in public space or self-managed places. In the self-narratives of the protagonists of these experiences, artists figure as "aesthetic operators". As noted by Enrico Crispolti (1977),

> the cultural operator, and the aesthetic operator in particular, in our case, is no longer in possession of a good to participate, to dispense; he is no longer a unilateral operator of culture, he does not, in short, impose his lesson, his action or his subjects: he is instead only a solicitor, provocateur, in a certain sense, of participation, and precisely of a participation that represents a moment of others' self-knowing growth; a participation in which the process (and possibly even the object) of aesthetic

5 The text is fully quoted in Verde, 2007.

> operativity is realized, overcoming the antithesis between creation (and research) and mediation. (p. 18)

It is inevitable to associate the maturation of this perspective with an atmosphere politically marked by strong decentralizing pushes and efforts for grassroots social participation, pursuing the systems of co-dependence most functional to the struggle. In those years, among the diverse range of extra-parliamentary political groups, the term *collective* occupied a prominent place: political collectives, self-consciousness collectives, feminist collectives emerged to represent distinct movements within broader organisational contexts.

One of the most renowned declarations of the Autonomous Collective of Painters of Porta Ticinese, formed in Milan in 1973, serves as a representative example: "The collective works against the capitalist organization of the State of labour and exploitation, rejecting the privileges of 'autonomy of the culture' and proposes as a group of militant operators the visualization of the contents of and the socialization of creative tools, recognizing that culture is that expressed by advanced struggles conducted by the proletariat" (Crispolti & De Grada, 1976).

The mobilization for a fairer world creates "impure identities" (many processes of subjectivation are as political as artistic, as social as cultural), in the name of an expanded and interdisciplinary creativity that seems to flesh out a *diffuse intellectuality*. Toni Negri (2014) appropriately captures the constructive drive that the 1970s made following the deconstructions of the 1960s: "From deconstruction our work consequently proceeds toward a collective process of self-valorization, of constructing circuits of value and signification completely autonomous, completely free from the market, definitively aware of the independence of desire" (p. 42).

3. 1980s and 1990s: Expanded Identities Through Artistic Acronyms and Innovations of the Net

Consistent with the overall climate of "reflux" that characterizes the 1980s, the artistic practices of the period stand out for a recovery of individual subjective dimension and a reappropriation of traditional languages[6] (such as painting), fundamentally rejecting the conception of art as a tool for political protest. To summarize with Lyotard (1982), "in the multifarious calls to suspend artistic experimentation, there is the same appeal to order, a desire for unity, identity, security, popularity [...]. There is an undeniable sign of this common disposition: the fact that for all these authors nothing is more urgent than settling the legacies of the avant-gardes" (p. 9).

Without going into specifics of the Postmodern debate[7], which has articulated the many shades and the intricated layers of the issue, I will limit myself here to pointing out how collective artistic experimentation suffered a setback at the dawn of a new decade, which was eager to move beyond the great ideological fervors of the previous one. Although they represent an exception to the widespread cultural atmosphere and often constitute themselves on the wave of imprinting left by previous militances, some significant collective projects took shape in the 1980s. Foremost, among these, I mention the Studio Azzurro group – born in Milan in 1982 on the initiative of Paolo Rosa, Fabio Cirifino, and Leonardo Sangiorgi – because one of its founders, Paolo Rosa, articulated the notion of the "plural artist" that I have used in this article. The operational and theoretical framework in which Studio Azzurro's plural practice developed originates from a close relationship with complex multimedia languages that necessarily required different creative skills and energies,

6 For a critical survey on the subject, see Buchloh, 1981.

7 For more in-depth coordinates, see Bauman, 1997; Harvey, 1991; Lyotard, 1984.

honoured by their collective signature[8]. "I must confront myself with a collective name, a collaborative practice that is that of Studio Azzurro", states Rosa (2017): "We have always argued, in these almost thirty years of our history, that it was right to put this nickname before the different roles, to express a creative climate indebted to many and diversified contributions" (p. 132). The Studio, aiming at the production of behaviors rather than things with which to suffocate a world already saturated with materiality, has over time developed researches on interactivity that in a certain sense extend the author's faculties to a spectator who has become a *spect-actor*, an active part of the work.

The trend of "return to order" after the easing of tensions that had swept the world from the mid-1960s certainly does not correspond to the recomposition of a presumed "lost identity", or a rediscovered individuality after the decline of the collectivist myth. The development of unbridled neoliberalism generates new forms of biopower from which the critical consciousness of many artists wants to disengage, while the crisis of the industrial Fordist paradigm opens up unprecedented scenarios that require the elaboration of new interpretive tools: creativity is increasingly subsumed within mechanisms of commodification, profit accumulation shifts from material to immaterial production, and precarity emerges as a generalized, structural, and existential condition – all elements that would later define the cognitive capitalism in which we still live today[9]. If forms of antagonism may survive, the pattern of conflict between big ideological blocs does not work anymore. Once again, the creation of collective artistic subjects provides insights for symbolic micro-subversions that challenge paradigms in the process of consolidation: it is the case of projects that do not present themselves as 'groups' or 'collectives', but rather as brands, commercial acronyms, or firms. Are born this way Group Material, General Idea, International Corporation, Information Fiction Publicité, Premiata Ditta, Banca di Oklahoma[10], whose critical appropriation of corporate aesthetics, advertising imaginary, and financial structures expose the underlying mechanisms of consumer culture.

Just to name a few examples, in 1983 Group Material proposed the project *Subculture*, in the form of an intervention in New York's subway system, in which the collective replaced the usual commercial advertisements inside subway trains with its own political and socially charged messages, using the same marketing language; meanwhile, in Canada, General Idea adopted a similar strategy with *FILE Megazine*, a satirical publication that mimicked the format of LIFE Magazine. Through this parody, which costed the artists a copyright infringement lawsuit, the collective highlighted how visual language and editorial structures shape public perception, and produce the commodification of culture. In Italy, the Bank of Oklahoma established itself as a fictional financial institution which produced banknotes, checks, and official-looking documents, parodying the growing dominance of financial markets and questioning the arbitrary nature of economic value.

The rhetoric of advanced capitalism and the language of marketing, increasingly pervasive due to the irreversible hyper-mediatization of the world, is replicated in a way that, beneath the sometimes parodic appearance, conveys a profound disquiet. The collective name here is not so much a space of common struggle as a force of depersonalization, reflecting the adaptation of creativity to the motor that regulates the world: the commodity. The individual finds their purpose in the collective, in the sense that identity appears to be

8 Think about the *sensitive environments* that the group has been proposing since '95. For further information, see Valentini, 2017; Mattei, 1999.

9 For an overview from a technical perspective, see Vercellone, 2006. For a political perspective, see Virno, 2002.

10 It is interesting to point out the presence, although I will not dwell on this, of projects carried out by individual artists that go in the same direction: to cite an example, the famous Ingold Airlines founded by the artist Res Ingold.

shaped by impersonal forces such as finance, globalization, and technology. By mimicking the production mechanisms prevailing in society as a whole, the international phenomenon of acronym-artists problematizes the order of values and behavioral patterns that, more or less consciously, inform our approach to consumer goods.

Emblematic are the considerations shared by the group BrigataEs, formed in Naples in 1992, in an interview with Silvia Bordini: after defining themselves as "homeless identities" and describing their work as a "self-mocking chronicle of an experience and at the same time the metaphor of a more general condition of change" (BrigataEs, 2002, p. 27), the group declares itself linked to an "aesthetic operation that, while ludically and critically underlining the loss of status of the contemporary artist, encourages the 'citizen' to recognize his own disidentity. The masked images thus become a sort of multiple mirrors that reflect the anonymity of the user" (p. 35). Even more programmatic is their statement:

> We aim for the recognition of those who show new possibilities of thinking, seeing, communicating. This is why we prefer to define ourselves as aesthetic operators and have chosen to present ourselves as a brand, an acronym. Our productions are not artworks but means of communication. Our target audience – which in part already exists but is mainly to be created – consists of 'aesthetic conscience objectors'. (p. 26)

Starting from the 1990s, with the evolution of the media society and especially with the advent of the Internet, the framework becomes more complex: while, on one hand, the Net represents a form of dispersion and dematerialization of resources, on the other hand, it serves as a laboratory for a new visual culture, offering spaces of interconnection between knowledges and disciplines, as well as new and extended possibilities for relationships. Expanded, multiple, fluid identities take shape within the Internet, imposing a rethinking of authorial logic itself, given that "communication as a directional transmission of information-culture from an 'author' to a 'user' no longer exists, having been dispersed in a network in which it is substantially impossible to trace all the channels (the pathway) and trace it back to a specific subject-author" (Pietroiusti, 1997, pp. 12-13). In taking charge of this undeniable reality, different forms of virtual communities[11] take root in a way that would have been unthinkable until recently, and that only partially find analogues in physical groups formed outside the Net. Various practices of dissemination and deterritorialization also include ecosystems that challenge the assumptions of the contemporary socio-political status through hacking, parasitism, and clandestinity, according to a sensibility that recalls the "artivist" militancy of the 1970s. Informal collectives arise from the grassroots using the Net as a free and anonymous space to pursue actions of displacement, "nuclei (cells) offering a magmatic matter in continuous dissipation (prodigality) and dispersion (dissemination)" (Perretta, 2002, p. 14): to name a few, think about Karen Eliot, Yomango, RTMark, etoy.com, The Yes Men, Las Agencias, Luther Blissett Project. The latter (which became Wu Ming in 2000) is one of the examples that gained visibility even in more mainstream circuits, even though the radical nature of its proposal has generated considerable scandals. In its 1995 manifesto, the group proclaimed:

> Capitalism dominates things and people by describing them: 'You are an I'. 'No, I no longer want to be an I, I want to be infinite Is!'. The collective name destroys the control mechanisms of bourgeois logic. Without the possibility of classification, power cannot

11 For further exploration, see Terranova, 2022.

> impose pre-cooked and pre-digested identities, nor operate to pit them against each other. (Luther Blissett, 1995)[12]

4. Conclusions

To conclude, it is useful to evoke some insights that Nicolas Bourriaud, curator, critic, and theorist of the so-called *Relational Aesthetics*, draws from the grafting of Felix Guattari's thought into the contemporary art field. The toolbox that the French psychoanalyst and philosopher provides to posterity primarily aims to denaturalize subjectivity, understood as a complex and not always pacified set of "plates", "planes", "machinery"[13], whose nature is not only individual, but also social and political. Paraphrasing Guattari, Bourriaud (2002) states that art

> is defined as a process of non-verbal semiotization, not as a separate category of global production [...]. The end purpose of subjectivity is nothing other than an individuation still to be won. Artistic practice forms a special terrain for this individuation, providing potential models for human existence in general [...]. There is nothing less natural than subjectivity. There is also nothing more constructed, formulated and worked on. (p. 88)[14]

Identity, far from being a natural attribute of the subject or an ontological *prius* to cling to, is thus a project to be imagined and constructed within vectors that transcend individual boundaries. Plural artistic practice, as a platform for the production of subjectivation, transposes onto the plane of authorship (and therefore makes explicit) the tensions with which each person constitutively negotiates the construction of his own identity: with technologies, with communication machines, with politics, with the social and intersubjective fabric in which one operates.

The fetishization of the author-demiurge and the romantic myth of creation, described by Roman filmmaker Anna Lajolo as "aggressive, authoritarian, vampiric", fade in favour of valuing the sharing flows that underlie every production process. Challenging a system that brands great personalities according to market demand, in which the signature of the individual author acts as "behavioural homogenization and reification" (Bourriaud, 2002, p. 94), the phenomenology of artistic groups, as it has taken shape since the second half of the 20th century, operationally restores the complexity of contemporary authorial identity. It rejects both the mythology of genius and the "death of the author" proclaimed by structuralists, instead proposing an idea of the artist as an "operator of meaning", who cultivates fertile ground for new phenomena of subjectivation while recognizing a multi-level interdependence: quoting Negri (2014), "art, as we have said, lives on production. Production lives on the collective" (p. 44).

REFERENCES

Alviani, G. (2013). Nuove Tendenze. Appunti e ricordi sull'arte cinetica da parte di un testimone e protagonista. In *Le Nuove Tendenze. Rivelazioni di un'arte percettiva e sensoriale,*

12 For in-depth studies and bibliographic materials on Luther Blissett, see https://www.lutherblissett.net/index_en.html (accessed on 12/06/2024).

13 On the imaginative language of Guattari, whose words often territorialize into images, see Bourriaud, 2002, pp. 86-87.

14 The text continues: "Guattari's contribution to aesthetics would be incomprehensible if we did not underline his effort to de-naturalize and deterritorialize subjectivity, expel it from his earmarked domain, the sacrosanct subject, and tackle the disconcerting shores with their proliferation of mechanistic devices and existential territories in the process of being formed [...]. We must thus learn to 'seize, enhance and reinvent' subjectivity, for otherwise we shall see it transformed into a rigid apparatus at the exclusive service of the powers that be" (Bourriaud, 2002, p. 89).

costruttiva e interattiva, 10 A.M. ART, Exhibition Catalogue. Retrieved from https://issuu.com/10a.m.artgallery/docs/le_nuove_tendenze;
Argan, G. C. (1961). Il Premio Lissone. In C. G. Argan et al. (Eds), *Catalogo della mostra XII Premio Lissone*, Exhibition Catalogue. Retrieved from flip.comune.lissone.mb.it/PremioLissoneEbook/Catalogo1961_1/;
Argan, G. C. (1963, September 21). Le ragioni del gruppo. *Il Messaggero;*
Balzola, A. & Rosa, P. (2011). *L'arte fuori di sé, Un manifesto per l'età post-tecnologica*. Milano: Feltrinelli;
Bauman, Z. (1997). *Postmodernity and Its Discontents*. Cambridge: Polity Press;
Bishop, C. (2012). *Artificial Hells. Participatory Art and the Politics of Spectatorship*. London-New York, NY: Verso;
Bourriaud, N. (2002). *Relational Aesthetics*. Dijon: Les Presses du réel;
BrigataEs (2002). Intervista. In S. Chiodini (Ed.), *BrigateEs. Artista volante non identificato*, Exhibition Catalogue. Ferrara: Sate Editore;
Buchloh, B. H. D. (1981). Figures of Authority, Ciphers of Regression: Notes on the Return of Representation in European Painting. *October*, 16, 39-68;
Crispolti, E. (1977). *Arti visive e partecipazione sociale. Da 'Volterra '73' alla Biennale 1976*. Bari: De Donato;
Crispolti, E. & De Grada, R. (Eds.) (1976). *La Biennale 1976/ Ambiente come sociale*. Venice: Edizioni La Biennale di Venezia;
Fontaine, C. (2020). *Human Strike and the Art of Creating Freedom*. Los Angeles: Semiotext(e);
Harvey, D. (1991). *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Hoboken, NJ: Wiley-Blackwell;
Kolb, R. & Richter, D. (Eds.) (2022). Documenta Fifteen – Aspects of Commoning in Curatorial and Artistic Practices. In *OnCurating*, 54[11];
Lyotard, J.-F. (1982). Intervento italiano. *Alfabeta*, 4(32);
Lyotard, J.-F. (1984). *The Postmodern Condition: A Report on Knowledge*. Minneapolis, MN: University of Minnesota Press;
Luther Blissett (1995). *Cosa vuole Luther Blissett*. Retrieved from https://www.lutherblissett.net/archive/335_it.html;
Mattei, M. G. (Ed.) (1999). *Interattività. Studio Azzurro, opere tra partecipazione e osservazione*. Perugia: Fondazione Umbria Spettacolo;
Meloni, L. (2020). *Le ragioni del gruppo. Un percorso tra gruppi, collettivi, sigle, comunità nell'arte in Italia dal 1945 al 2000.* Milano: Postmedia;
Negri, T. (2014). *Arte e Multitudo*. Roma: DeriveApprodi;
Perretta, G. (2002). *Art. Comm., Collettivi, reti, gruppi diffusi, comunità acefale nella pratica dell'arte: oltre la soggettività singolare*. Roma: Castelvecchi;
Remotti, F. (2010). *L'ossessione identitaria*. Roma-Bari: Laterza;
Rosa, P. (2017). In V. Valentini (Ed.), *Studio Azzurro. L'esperienza delle immagini*. Milano-Udine: Mimesis;
Terranova, T. (2022). *After the Internet. Digital Networks between Capital and the Common*. Los Angeles, CA: Semiotext(e);
Verde, G. (2007). *Artivismo tecnologico. Scritti e interviste su arte, politica, teatro e tecnologie*. Pisa: BFS;
Vercellone, C. (Ed.) (2006). *Capitalismo cognitivo. Conoscenza e finanza nell'epoca postfordista*. Roma: Manifestolibri;
Virno, P. (2004). *Grammar of the Multitude and Multitude between Innovation and Negation*. Los Angeles, CA: Semiotext(e).

Links
https://dis.art/
https://documenta-fifteen.de/en/lumbung/

SECTION

SECTION 5

FREE CONTRIBUTIONS

FEDERICA MADONNA
Università degli studi di Cassino e del Lazio Meridionale
f.madonna@unicas.it

THE OVERLAP OF MINDREADING AND METACOGNITION: AN ELIMINATIVIST EXPLANATION OF COGNITIVE EMPATHY[1]

abstract

This paper suggests that the concept of cognitive empathy, as currently presented in much of the literature, may rest on conceptually fragile assumptions. The argument is developed in three steps: a reconstruction of the historical misunderstandings behind the idea of empathy; a comparison between mindreading and metacognition; and a proposal for an alternative view of intersubjectivity based on what is defined here as a "relational void". Rather than demonstrating definitive claims, this contribution aims to open a critical reflection on the theoretical status of cognitive empathy and its possible redundancy when viewed through the lens of metacognitive processes.

keywords

consciousness, cognitive empathy, mindreading, metacognition, relational void

1 I thank Prof. Andrea Lavazza for the valuable suggestions he provided me during the writing of this article.

Phenomenology and Mind, n. 29 - 2025, pp. 176-194
DOI: 10.17454/pam-2902
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In recent decades there has been much talk about *empathy*; at times it has been placed in a neurobiological context; others in a relational context; others yet in a cognitive context. It is no coincidence that attempting to provide a *clear and distinct* definition of the phenomenon is fraught with difficulty, as there are as many definitions as there are scholars engaged in the debate[1].

Indeed, the term "empathy" has undergone a remarkable semantic evolution, resulting in a series of frequently divergent definitions. Some authors have given priority to its affective aspect[2], describing it as the ability to resound with the other on an emotional level. Others have emphasized its cognitive dimension, understood as the ability to take another's perspective or infer mental states[3]. In addition, more recent contributions have sought to integrate neurobiological, social and moral components, emphasizing the adaptive function of empathy[4]. While this theoretical plurality enriches the debate, it also runs the risk of creating conceptual confusion, especially if a definition is assumed to be universally agreed upon. It is precisely this complexity that calls for a critical re-examination of the concept of cognitive empathy as it is commonly understood.

What holds this conceptual tangle together is, arguably, the relationship between two individuals (henceforth: an I relating to a You) which is capable, through this strange phenomenon, of "coming into contact". The "strangeness" could be explained in a very haphazard mixture of aspects that, in the very history of the concept, have become intertwined without any longer providing conceptual clarity as to what these very aspects were supposed to explain. We refer to the fact that the very sketchy explanation of empathy as "the ability of the I to experience first-hand the emotions of the You," the result of which would be that the I and the You feel the same suffering or joy conveyed by the interlocutor at that moment, has been enriched, in a symbolic way from the 1990s onward, with neuroscientific aspects (in this regard, mirror neurons represent the most relevant aspect) to explain this particular inter-relationship; the consequence was that from *empathy* we have ended up talking about *mentalization* or the ability to "read [neurobiologically, Ed.] the mind of

1 Cf. Decety & Ickes, 2011.
2 Cf. Eisenberg, 2018; Aaltola, 2014.
3 Cf. Baron-Cohen, 2025; Decety & Ickes, 2009.
4 Cf. Cuff, Brown, Taylor, & Howat, 2016; Cameron, Conway, & Scheffer, 2022.

the You"[5]. This distinguishes it from empathy, which is understood as an emotion felt only by the I, regardless of the situation, the context, and whether the interlocutor is present or not[6].

This "evolution" has occurred in stages so much so that, following Martin Hoffman's[7] timely analysis, in the current debate the "empathy problem" seems to have found its own universally shared subdivision: on the one hand, basic empathy or primary aspect; on the other, cognitive empathy or secondary aspect and its possible relationship to how morals came to exist. In the first case, referring to the relentless development of cognitive neuroscience in recent decades, it is understood as the set of all the neurobiological mechanisms underlying the phenomenon (experiencing first-hand the emotions of the other on oneself"); in the second, emphasis is placed on a different and founding aspect of the phenomenon, which involves the cognitive effort of the I to understand the experience which the You communicates to it, which in itself has the effect of experiencing first-hand the emotions of the You; this is referred to as secondary or cognitive empathy. An outgrowth of this aspect is the interest that has arisen around the idea that there may be a relationship between the role of empathy and the emergence of morals[8], extending the question of whether altruism can identify a pre-social form of empathy, whether they are two distinct phenomena, or whether one is the cause of the other.

While basic empathy, therefore, has initiated a field of research aimed at understanding which areas of the brain are involved in giving rise to empathy itself, cognitive empathy is the subject of numerous interpretations, which focus on how this cognitive effort of the I can take place in order to understand the other, the result of which is a cognitive form capable of "entering" into the interlocutor's mind and innermost experience. The characteristics of cognitive empathy can be summarized in three: the processing time of the message; the cognitive effort to decode it; and, lastly, the psychic distance between the two interlocutors[9].

As can be seen, in the cognitive dimension, the interpretation of the empathic phenomenon is completely separated from the neuroscientific aspect and, while up to the early 2000s this "field of research" focused on the different facets of the phenomenon itself, identifying its different characteristics and nuances, in the last twenty years (thanks to the continuous development of neuroscience) the field of research is not "clear", uniting both the primary and secondary aspects. This fusion has given rise to a real phenomenon in itself, *mentalization*, which allows one of the two interlocutors to "enter into the mind of the other, reading its intentions, desires and beliefs"[10] through specific neurobiological mechanisms for mindreading"[11], and then recreate them-neurobiologically and cognitively-into oneself, opening up perspectives on how to "project towards oneself" a thought which originated in the other's mind and "transport" it into one's own (the problem of metacognition).

Thus, mindreading represents the most recent evolution of the "empathy phenomenon" to date, suggesting that the multiple successive interpretations of the "I-You" relationship suggest that the concept of "cognitive empathy" itself is structurally fragile, opening up the possibility of its critical revision.

In fact, the intention of this article is precisely to suggest that a way forward in this direction is possible by critically exploring alternatives that would lead one to think that some

5 Cf. Almeida et al., 2024.
6 Cf. Aaltola, 2014.
7 Cf. Hoffman, 2010.
8 Cf. Cameron, Conway, & Scheffer, 2022.
9 Hoffman, 2010.
10 Cf. Stietz, Jauk, Krach, & Kanske, 2019.
11 Hutto et al., 2011.

interpretive models used to explain the phenomenon (e.g., those that have relied on the use of folk psychology or the use of the concept of mind) may have contributed to the construction of illusory representations about the role and explanation of what cognitive empathy[12] is; this could lead to the delineation of different models of the mental that, in a likely explanation, lead to the introduction of the concept of "relational void".

In order to achieve this goal, a number of key steps will be necessary. In particular, the first section will explain the emergence and revival of the concept of empathy up to its declination as "cognitive empathy", and the second section will focus on the transformation of the concept of cognitive empathy into "mind reading", clarifying the reasons why a discussion has started in the contemporary debate on how it is possible to "think about one's thoughts (metacognition).

This is done in order to provide the conceptual premises to address the ultimate purpose of this paper, which is to suggest that it is plausible to initiate a critical revision of the concept of cognitive empathy itself.

It is important to clarify that this paper does not intend to question every possible meaning or declination of cognitive empathy. The analysis focuses specifically on certain theoretical formulations which currently prevail, such as that of Martin Hoffman, which we believe contain elements of ambiguity or conceptual redundancy. We do not rule out the possibility that other approaches – especially phenomenological and embodied ones – may offer more coherent readings of the phenomenon, but they are not the specific focus of this discussion[13].

Thus, in the next section, we will hypothesize that mentalization is a conceptual duplication of the phenomenon of metacognition theorized by John Flavell in the 1980s, since it overlaps perfectly with "mind reading," in order to draw the reader's attention in the fourth section to a naturalistic reading of the mind, anticipating what is meant by "relational void," i.e., an alternative explanation of intersubjectivity that will be made explicit only in the conclusion.

2. The origin of the misunderstanding: from *sympathy* to mindreading

The origin of the concept of cognitive empathy, as it has been defined, is relatively recent, since it has been included in scientific debate only since the 1950s, following a revival of the concept by philosophers Willard van Orman Quine, Hilary Putnam and Robert Nozick. Subsequently, it underwent considerable changes until its final transformation into "mindreading".

Understanding these changes, starting from their historical origin, involves having a clear idea of the common thread that has formed the trail on which various misunderstandings and various answers have been provided in order to address old dilemmas, attempting to completely abandon any recourse to metaphysics. Such claims are reflected in the different steps in the evolution of the phenomenon, beginning precisely with David Hume, who, both in his *Treatise of Human Nature*[14] and in his *Enquiry Concerning the Principles of Morals*[15], explained

12 Cf. Cuff, Brown, Taylor, & Howat, 2016.

13 The reference to Hoffman is understood here as representative of a widespread theoretical line, particularly in the fields of developmental psychology and cognitive neuroscience. It is obvious that the position advocated in this paper does not claim to apply to every notion of cognitive empathy, but merely specifically criticizes this theoretical approach and its developments with respect to mindreading and meta-representation.

14 Cf. Hume, 2023b. We point out that the literature on the interpretation of *sympathy* in Hume offers two explanations: one argues that the latter can correspond to a kind of "emotional fusion" between different individuals who may not live in the same social context; the other rules out such an interpretation, since in both the *Treatise* and the *Enquiry* Hume precisely distinguishes sympathy/natural benevolence from emotional contagion, pointing out that these are two different aspects. Here we agree with the latter interpretation.

15 Cf. Hume, 2023a.

that the foundation of morals consisted of *sympathy* first, which is then transformed into *natural benevolence*.

In the first work, the Scottish philosopher claimed that *sympathy*, in addition to defining the mutual action of things which affect one another, is able to make us *enter so deep into the opinions and affections of others whenever we discover them*[16] precisely because it represents the internal principle of the passions by which the transition from *ideas to impressions* is achieved. An epistemological tool that, excluding any "emotional fusion" between interlocutors, originates in itself and results in *imaginative participation in the lives of others*[17], providing a "primary" knowledge of the other. In *Enquiry* these aspects of *sympathy* are extended by Hume himself to a universal explanation as to why in the human there is a feeling of benevolence, of "openness to intersubjectivity", which explained why two or more subjects are able to cooperate, converse, and interact.

It is not a methodological error to start precisely from the Scottish philosopher in order to understand the origin of today's cognitive empathy, since in his philosophy there are several points of junction with modern conceptions of empathy, both in its relation to morals and in the relationship between two or more individuals. The latter aspect, completely secondary for Hume as in the entire current of the eighteenth century, was given special attention only as a result of the interpretation of the meaning of *ein-pathos* ("entering into the passions") deduced from the philosophy of Friedrich Schleiermacher who, by conveying the emotional aspect of sympathy to the field of art, argued that in all works of art the artist transposed a part of himself and his "authentic" I onto the canvas, therefore the viewer "knew" and "emotionally felt" the artist's interiority. Only as a result of such a "change of direction" did the intersubjective aspect enter the collocation of the concept of empathy which, transformed completely from Hume's real intentions, became, both with Theodor Lipps and Carl Rogers, an explanation of why one understands the other by experiencing first-hand the interlocutor's experience.

Also in this respect there has been a controversial methodological and interpretative approach on the evolutionary steps of the phenomenon, as today's debate recognizes Lipps himself as the "founder of empathy"; upon a careful reading of his works, it is noted that the aim that moved the philosopher from Leipzig was not to provide an explanation of intersubjectivity, but to "give order to the world" as Immanuel Kant himself, through the categories, did in his *Critique*.

Sharing the views of contemporary philosopher Dan Zahavi, Lipps seems to explain empathy as the result of a motor mechanism rather than as an explanation of the mental life of others[18]. Following the criticism of Edith Stein or Edmund Husserl, the German philosopher sees a discrepancy between the phenomenon to be explained (empathy) and the phenomenon actually explained, because to resort to the "commonality of feeling" between the I and its interlocutor simply by claiming that it is due to an "emotional resonance" – as Lipps claims – is not to have explained the difference between recognizing another's emotion and knowing it. They are two completely different epistemic aspects, says Dan Zahavi, who argues that since empathy is a phenomenon that can be interpreted in different ways, to the point of being almost a polysemantic term, it would be useful to always take into account in its use the phenomenological difference between knowing others, oneself and external objects, so that

16 Hume, 2023b, pp. 334-335.
17 Cappiello, 1983, p. 87.
18 Cf. Zahavi, 2014.

we try to keep the different levels of explanation of the empathic phenomenon separate[19], although we may wonder if it would not be more honest to abandon the use of the term altogether[20].

Continuing with the analysis of the concept, a further transformative step of cognitive empathy derives precisely from analytic philosophy, especially from Quine who, through the use of "radical translation," with the precise intention of "overcoming" Carl Gustave Hempel's Deductive Nomological model by criticizing it, recommended giving philosophical thought to the concept of empathy, understood as a real cognitive tool capable of providing knowledge on the desires or volitions of the other. This operation of restoring the concept was endorsed by Putnam and Nozick themselves: the former by providing *simplicity and plausibility*[21] ; the latter, by paving the way, following the publication of the paper by the two ethologists David Premack and Guy Woodruff [22], for what would become "mindreading".

In such illustrative passages, a gradual transformation of the concept of empathy is evident: from *sympathy* to mentalization, suggesting the interpretation that this phenomenon responds to questions left unresolved in modern times, that is when, symbolically from Descartes, the examination of the problem of the nature of the mental, whose reflections are still debated today. An example of this is Gustave Hempel's Nomological-Deductive model, which – by the very nature of the analytical approach – completely excluded the metaphysical and emotional aspects from the understanding of the phenomena since both were unable to respond to the rational formalism of syllogism. In this context, while, on the one hand, Quine had the insight to "break out" of this schematism, restoring the concept of empathy, on the other hand, the latter was useful to Quine himself as a purely natural element to explain the relational aspects of the mental experiment.

In this way, by providing empathy with the nature of an emotion, and placing it in the realm of the pragmatist, Quine was able to respond to Cartesian dualism (are we thought or matter or both?) by not resorting to any speculative metaphysics on the *cogito: res cogitans*, that is, it was naturalized. It is along these interpretative lines that, subsequently, in order to avoid any form of mentalism, cognitive empathy was transformed from an "embodied instrument" of the person into an "objectified instrument" with Wilfrid Sellars' theorization of the "Myth of Jones," first, and of folk psychology, later. A conceptual framework that, by resorting to what is most "existential" and pragmatic, is explained as:

> all the platitudes you can think of regarding the causal relationship of mental states, sensory stimuli and motor responses [...].
> And also, all the platitudes to the effect that one mental state falls under another [...]. Perhaps there are platitudes of other forms as well. Include only platitudes which are common knowledge among us: everyone knows them, everyone knows that everyone else knows them, and so on (Lewis, 1972, p. 256),

transforming cognitive empathy into a valid stratagem to avoid the question of the nature of the mental.

Despite a controversial conceptual history, empathy has provided the current debate with the incipit of several fields of study, which can be summarized in five major directions:

19 Zahavi, 2001.
20 Ivi.
21 Cf. Putnam, 1995.
22 Cf. Premack & Woodruff, 1978.

- *Temperamental or personality variables related to empathy-related responding*, in which attention is paid to the degree of empathic intensity experienced by the agent placed before a victim's distress, whose variables are their temperament and personality;
- *The development of empathy-related responding*; in fact, for the past few years, studies aimed at the study of the emotional-empathy development of minors have arisen from their early stages of cognitive development;
- *The relation of empathy-related reactions to social behavior, including prosocial behavior, aggression and social competence*; taking up and reproducing Milgram's experiment, attention has been paid to how the perpetrator's empathy is annihilated before the authority figure or how, on the other hand, empathy may be the motivation for the agent to engage in prosocial behavior, perhaps at the basis of altruism[23];
- *Gender differences in empathy-related reactions*, in whose research an attempt is made to provide an explanation of why one gender would be more inclined to feel empathy than the other;
- *Socialization correlates*, in which the limitations to which empathy is subjected are highlighted, referring, in particular, to the different response of the agent when faced with a victim's situation of suffering: from a close family member or relative to a complete stranger.

The latter, stigmatized (in a few lines here) by Nancy Eisenberg[24], summarize the heart of the problem[25]: despite the fact that it is almost impossible to define empathy unanimously, an attempt is made to make the entire theoretical construction of the phenomenon plausible by making the explanation of what it is plausibly true.

3. From cognitive empathy to mentalization

It has already been said that the concept of empathy has undergone several shifts in meaning from Hume's original intent to contemporary times, and it has also been mentioned that in today's debate, as a result of the ongoing neuroscientific developments of the last thirty years, it has been agreed to separate the meaning of empathy from a basic to a cognitive interpretation; a "separation" that, if for nearly twenty years it was the pragmatic basis of a certain explanation of cognitive empathy (the "embodied simulation" of folk psychology), in recent years this – unanimously recognized – distinction is constantly "betrayed" in an ongoing attempt to provide simplicity and plausibility to the phenomenon itself. This suggests that the main problem of cognitive empathy is based precisely on the network of cascading processes that it sets in motion to ensure that the steps related to "understanding and feeling the emotional states of others" are followed; which would make intersubjectivity possible: "entering into the mind of the other" in order to acquire an "epistemic" knowledge of it; "recreating it in oneself"; experiencing the same emotion of the other's subjectivity cognitively and sensorially.

An example of a "basic" approach to explaining empathy is that of Frederique De Vignemont and Tania Singer, who argue that empathy could be defined as a "modular", unconscious and implicit response to the social situations in which the individual interacts. Sharing the same neural mechanisms, the empathic response is activated with the help of two different modular pathways, each with a specific function: a first mechanism involves a conscious, voluntary use of the empathic response, taking advantage of the subject's conscious

23 Krebs, 1975.
24 Lewis, Haviland-Jones, 2018, pp. 677-691.
25 Wispè, 1986.

control of the affective response to the other's emotion; a second mechanism involves an implicit, unconscious, immediate assessment of the other's emotional situation, capable of influencing the emotional strength of the response. In both situations, according to the two authors, the empathic response is activated directly and automatically, emphasizing that the latter has a purely social and epistemological role: that is, it is able to provide a direct estimate of the possible future actions taken by the interlocutor (social role), giving an explanation (epistemological role). A limitation of such an instrument, also according to the authors, is the sharing of the same experiential and cultural background[26].

What is evident in De Vignemont and Singer's explanation of empathy is certainly the value of the constantly developing neuroscientific research that attempts to identify the brain areas or mechanisms involved in the phenomenon; what does not seem to be the object of analysis, however, is an "ontological" explanation of empathy itself: in fact, the two authors start from the assumption that the phenomenon exists and that it can have certain or certain other explanations.

Alternative approaches to intersubjectivity – such as phenomenological[27] or direct perception approaches[28] – offer different interpretations of the relationship between self and other that are not based on simulation or meta-representational models. This paper, however, takes a different perspective, aiming to explore the conceptual fragility of the notion of cognitive empathy as it is commonly understood. This does not mean that such approaches do not deserve careful consideration in relation to the themes of this article, which can be done in a subsequent in-depth study.

It should be acknowledged, however, that alternative approaches to the theory of mind – particularly those of direct perception – have proposed different models for explaining intersubjectivity, freeing themselves from both simulation and meta-representation. Authors such as Shaun Gallagher[29], Pierre Jacob[30] and Joel Krueger[31] argue that understanding the other can occur through direct, embodied and contextually situated access, thus reducing the role of "internal" and inferential mental processes. Although this paper does not explore these perspectives in depth, they pose a relevant challenge to the automatic identification of empathy with complex cognitive operations and deserve careful consideration in future work, especially in relation to the critique of mentalization as a conceptual duplicate of metacognition.

Continuing the analysis, the moment when cognitive empathy symbolically became "the ability to enter the mind of the other to interpret their intentions, beliefs and desires" took place, as said above, with the publication of the paper by the two ethologists Premack and Woodruff in 1978, in which the two authors questioned whether chimpanzees also had an ability to understand the intentions of others, consequently adopting a behavior appropriate to the stimulus understood.

On account of the peculiarities of the task implied by this ability, people began to speak about a real *theory of the mind*, which could be applied to nonhuman animals as well. Hence the origin of *mentalization*, i.e., a further transformation of the common meaning of cognitive empathy to indicate, specifically, a particular ability, neurobiologically inherent in humans and not only in the human species, to *actually* "enter" the mind of the other; from a "mere"

26 Cf. De Vignemont, & Singer, 2006.
27 Cf. Zahavi 2001.
28 Cf. Gallagher, 2008.
29 Cf. ivi.
30 Cf. Jacop, 2011.
31 Cf. Krueger, 2018.

cognitive effort that allowed the I to "understand" the You, we moved on to the identification of a "functional module" of the entire brain capable of *letting one's mind enter* that of the other.

From a strictly speculative point of view, the originality of the two ethologists' essay was to fit into the progressive construction of the *empathy stratagem* to solve the problem of intersubjectivity; that is, there was a need for an element that was able to explain why an agent, situated within an already given context (*folk psychology*), was able to "understand the other". The use of means which were already known (shared beliefs) was not sufficient, as one sought to provide a rational justification for the kind of epistemic knowledge that empathy could provide; without "scientific" support, the latter simply appeared to be mere intuition. Therefore, Premack and Woodruff's paper provided the "lifeline" from the collapse of the *Folk Psychology* theory and its countless internal facets, which can be summarized in the two great families of *Theory of the Theory* (TT) and *Simulation Theory* (ST).

TT refers to the explanation of intersubjectivity by using a set of "common laws" (*folk*) through which beliefs, desires and volitions are attributed to the other than Self: *folk Psychology*, therefore, is equivalent to an "already given" theoretical framework used by the agent through which he can understand that of the interlocutor[32].

In the case of ST, on the other hand, "common sense psychology" is interpreted as a "false theory", referring to completely non-existent objects (mental states); therefore, the explanation of intersubjectivity is completely entrusted to the brain processes of the agent himself, which are able to initiate "internally" – in a "simulated" manner – the "understanding" of others' mental states thanks to special cognitive mechanisms which, from the 1990s onward, were hypothesized to be mirror neurons[33] along with other areas of the entire encephalon[34].

Although the simulation theory has been strongly advocated by many to explain the way in which we "understand" or interact with each other, the "ST paradigm" has been in crisis for several years due to severe criticisms of the way it works. One of them is the one put forward and shared by Shaun Gallagaher[35], according to which ST is based exclusively on the neural process (high or low level) by which the I, drawing on its own experiential background, "reproduces in itself" the emotional state of the other, experiencing the same feeling.

This means, Gallagher continues, that empathic understanding of the other is always limited only to what the I has experienced first-hand, thus excluding from the simulation all situations outside its experience; how, then, would intersubjective openness to the other be explained if such understanding were limited only to what the I knows first-hand? How, then, would openness to the other be explained?

According to Gallagher, such an objection to ST renders the entire paradigm inadequate to explain the empathic phenomenon, also for the reason that it would always be limited to the experiential background of the empathizing subject; on the other hand, what could help explain empathy itself, according to Gallagher, is the observation that understanding the other is always based on one's own model of understanding and knowledge of the context in which the other acts and operates. In short, empathy, in order to exist, requires a wealth of narratives, what the author himself calls the "hermeneutic background of empathy," which provides the I with an understanding of the context in which the You acts and, consequently,

32 Cf. Stueber, 2010.
33 Cf. Rizzolatti & Sinigaglia, 2006.
34 Cf. Baron-Cohen, 2025.
35 Cf. Gallagher, 2012.

what it might feel in that given situation. The ST lacks an explanation of how such a "context of understanding" or its role in the empathic phenomenon is obtained.

Although Gallagher's explanation focuses on a relevant aspect of empathy, namely the finding that there is an experiential framework that comes into play when "empathy is experienced," it fits into a heavily debated strand of studies on the "natural" relationship between empathy and altruism; the problem is to define whether the one is the prosocial behavior at the origin of the other, or vice versa. The aspect that is not shared here, because of the very intention of the thesis in question, is the starting assumption that the empathic phenomenon requires a narrative context already given to the I; this means accepting the indisputability of the phenomenon itself, leaving aside the strong conceptual limitations it presents, its fruition, for example, in subjects who are spatially and physically distant[36]; how, then, could we use empathy in order to "understand" the affective situation of the other when, in such a condition, we would not know the history or the context in which it is found? Moreover, the knowledge of the "already given" context calls to mind folk psychology, and thus does not move far from the structural problems that the latter poses.

It should be noted that the introduction of common sense psychology in the debate has led to a further change in the definition of "cognitive empathy" which, instead of focusing on the nature and epistemic justification of intersubjectivity, "reinvented" the problem by "including" it into the I itself: what kind of neurobiologically structured structure does the I possess to "enter into relationship" with the other? Or, rather, to "enter," through its thinking, into the mind of the You in order to know and interpret its intentions, beliefs, and desires in order to then "bring back into itself" the result of this shift outside the body, created by the mind of the I itself?

The problem regarding mentalization or mindreading took shape precisely starting from this question so generically summarized and, in this regard, the pillar of an entire debate is a target article by Peter Carruthers, *The relationship between mindreading and metacognition*[37] – published in 2009 – to which he is credited with both collecting and discussing the various arguments made in support of four generic models of explaining mentalization (*Model 1: Two mechanism, two modes of access; Model 2: One mechanism, two modes of access; Model 3: Metacognition is prior; Model 4: Mindreading is prior*) and with providing one that clarified the relationship between the mindreading structure directed toward the other and the "return into Self" of this interpretation.

The central point of Carruthers' position, which is helping to shift, once again, the focus of the debate on cognitive empathy, is the consideration that humans are "avid mind readers"[38]: they attempt to attribute to the other beliefs, thoughts, states or volitions experienced at that moment as an integral part of their own actions, triggering a process of "meta-representation", i.e., a process in which the object of study can become either the other than Self (therefore, we would be within the "classical" problem of "cognitive empathy") or toward Self; in the latter case a process of metacognition is set in motion, that is, the effort or process of thinking about the same "mind reading" thinking process, which results in an equal process of cognition, but directed toward oneself.

The questions that are following are inherent to the kind of relationship subsisting

36 Cf. M. Hoffman, *Empathy*, op.cit.

37 Cf. P. Carruthers, *How we know our minds: The relationship between mindreading and metacognition*, in "Behavioral and Brian Sciences", 32, 2009, pp. 121-138.

38 Id., *How we know our minds*, op. cit., p. 121.

between the former and the latter process, posing the question of whether there is a difference (if so, what kind?) between access to other minds and access directed toward Self.

In this respect, in the mindreading/metacognition relationship several protagonists in the debate have accused Carruthers of eliminating the problem of the I's access to itself: this would no longer imply a relationship between reading the mental states of others and the underlying cognitive capacities, but a dualistic explanation of the relationship. How, in fact, is the I able to act, itself, as an object of study? And what "structures" would be involved: only the "neurobiological" ones, naturalizing the mental, or only those of the mind, distinct from those of the brain?

It is no coincidence that many advocates of the position advanced by Carruthers have suggested that he focus only on the I-other relationship. In fact, the way the US philosopher's proposal has been structured would suggest a distinction between the mechanisms deputed to mindreading and those deputed to the interpretation of metacognitive states almost as if they were two distinct substances, one of which has already been inserted into the process and the other "to be inserted" at a later time[39].

In light of what has been said so far, it would be useful to include an outline of the steps elaborated and discussed in order to highlight the controversial aspects of the "empathy phenomenon" that would suggest conceptual limitations of both cognitive empathy and the mentalization process:

1. Transformation of the meaning of *sympathy* in Hume to "empathy," understood as *einpathos*, in both the artistic (Schleiermacher) and theoretical (Lipps) currents, assigning it the "ability to feel, first-hand, the emotions of others" and to provide knowledge;
2. Quine's revival of the concept in the "radical translation" experiment, assigning empathy an epistemological use;
3. Putnam and Nozick's specification of the characteristics of *simplicity and plausibility* and knowledge of others' behavior to the concept of empathy;
4. The risk of providing a metaphysical status to empathy and, thus, the fear of slipping into the dilemmas of Cartesian dualism, provide the *incipit* for seeking a transcendental and pragmatic nature to empathy; in particular, to the nature of the mental effort involved in the same capacity. Wilfrid Sellars, with his theorization of the "Myth of Jones," provides the solution: the theorization of common sense psychology, namely, a "background theory already given and known," within which one objectifies the set of desires, beliefs and volitions that the You might feel (the mind of any member of the given community is objectified) and uses empathy as a mere search tool of the I for a suitable feeling in a given situation in the set of "already-given knowledge" provided by *folk psychology*;
5. The collapse of the two great families of common-sense psychology (the ST and TT) has shown its vulnerability, having given cognitive empathy only an intuitive explanation of its nature and not a theory based on its truth. In this regard, Premack and Woodruff's paper provides a solution, by further shifting the problem of the empathic nature: the inclusion of a theory of the mind in the explanation of cognitive empathy. In this way, the empathic capacity, from objectified tool in *folk psychology* "goes back to being" the "interiority" of the I in the theory of the mind;
6. The debate questions how, through a theory of the mind, the I is able to "understand" that of the You; Peter Carruthers' essay symbolically initiates the explanation of the

39 Cf. Dennett, Huebner, 2009.

process of mentalization, namely the ability to "read the mind" of others and "enter it" to understand their desires, beliefs and volitions. The division between basic empathy and cognitive empathy is betrayed and there is an attempt to explain the latter through recourse to the former.

7. The last frontier of totally sliding from the original intent of the meaning of *sympathy* occurs by the emergence of new questions concerning the "neurobiological capacity" of mindreading: how can we explain the kind of knowledge that the "mentalization module" can give the I if it were to address the capacity of mindreading to itself and no longer to the other?

This last question, it has been said, transforms the debate on cognitive empathy once again, raising the question of the kind of relationship there is between metacognition and mentalization, but on closer analysis it will become clear that this apparent novelty is actually the recovery of a phenomenon already theorized in the 1980s by John Flavell, metacognition which, in a forward-thinking manner, already explained decades in advance what the invention of mindreading denotes.

To this end, to suggest to the reader that such an interpretation will be necessary, a brief analysis of both what Flavell had meant by metacognition and reasons why it can be overlapping on the phenomenon of mindreading.

4. Metacognition and the overlap with the mentalization process

The discussion on metacognition in relation to mindreading is not an absolute novelty in the debate on metacognition itself, but it is in light of its transformation over time of the concept of cognitive empathy.

Upon a careful analysis of metacognition, both in its historical genealogy and in its ongoing development, we can see that as early as the 1960s it has made its way into the debate on emotions and the kind of knowledge they involve.

To be precise, U.S. psychologist Josef Hart[40] had identified the meaning of "metacognition" as a "feeling/perception of knowing"[41] in the light of his research, concerning the functioning of memory and how the person was able to monitor and control the set of information stored. He defined "metamemory" as the ability to establish the predictive reliability of the "feeling of knowing" in the evaluation of one's mnemonic abilities.

The term "meta," at least in Hart and in later studies inspired by him, referred to the subdivision of mnemonic ability: a first task related to remembering information and data; the second to monitoring and "calling to consciousness" such information, and perceiving this recollection through the "feeling of knowing".

While Hart is the symbolic leader of the debate on metacognition, influencing and stimulating different fields of research or investigation within it, we owe the most common and most discussed meaning of metacognition to an article published by John Flavell. As mentioned above, the American psychologist published a very short article in 1979, entitled *Metacognition and Cognitive Monitoring. A New Area of Cognitive-Developmental Inquiry*[42], in which, building on Hart's analysis, he argued that metacognition indicated knowledge and cognition about cognitive phenomena, that is, about all processes (cognitive, precisely) such as

40 Cf. Hart, 1965.

41 In the current debate, this is the reason why we are focusing on the "tip of the tongue" phenomenon to demonstrate the critical role of metacognitive processes in the development of memory and, more generally, the issues involved in metacognition. For further discussion, see Heyes et al., 2020; Rouault et al., 2018.

42 Flavell, 1979.

attention, memory, problem solving, self-control and self-awareness along with all social and cognitive processes of which an individual was capable.

The "field of investigation" of metacognition expanded considerably, no longer relegated only to the processes of memory control and monitoring, but to the totality of all the "knowledge" that the person could obtain, by turning their attention, memory, thinking and cognition, in general, toward themselves. From this point of view, the "meta" of the process identified by Flavell no longer stopped at the mere distinction of roles and capacities of mnemonic tasks – as in Hart – but indicated the possession of a theoretical knowledge of what the subject knows, understands, and remembers and the related processes (what, when and why). Therefore, a kind of "subdivision of levels" between the thought of an I, looking outward or towards a You, and a more complex thought looking toward oneself; what many years later would become "mind reading".

In the first few paragraphs of the article, the American psychologist endorses the difference between *metacognitive knowledge*, which, as noted, is the result of the process of self-questioning by the Self, and *metacognitive* experience, overlapping with Hart's "feeling of knowing," arguing that both can be considered almost the same phenomenon: one influences the other and vice versa, precisely because the activation of purposes or strategies of action for the achievement of an end (the other two classes of the phenomena he identified) affect both access to memory and the awareness of one's cognition: metacognitive experience is the former; metacognitive knowledge is the latter. Therefore, since metacognitive knowledge has no different nature than other knowledge "stored" in one's memory, according to Flavell it can also be defined as such when it influences the course of action taken by an agent for a purpose in an unconscious manner; this means that awareness or consciousness does not play a relevant role in "self-mastery" and that it "simply" represents an option.

Beyond this controversial aspect, it should be noted that if we were to provide a merely phenomenological explanation of metacognition, it could be outlined in this way:

- X conceptually represents mental states;
- X conceptually represents [others', Ed.] mental states with their intentional content toward which they are "directed" [...];
- X understands, through certain meanings, the relationship between an agent's mental states and their respective behavior [...]. This understanding enables X to make predictions about others' behaviors and to explain their behaviors[43].

Paraphrasing the above, in the first aspect we could refer either to cognitive knowledge or the metacognitive experience: an agent may or may not be aware of their cognition and have first-hand experience of it; in the second, by using the three sub-categories identified by Flavell as pertinent to the nature of metacognitive knowledge (people, tasks or strategies), it is explained that in the inter-relationship with the other, the cognitive effort to "understand" the You does not identify a specific ability to go beyond oneself to "simulate," "reproduce," or "change one's point of view" to produce knowledge about the other, but a self-directed ability

43 Hutto et al., 2011, p. 376.
We read: *The standard view is that an agent X engages in mindreading only if:*
1. *X conceptually represents mental states (e.g., beliefs, desires and perceptions).*
2. *X represents mental states with their intentional (with a "t") content, i.e., that which they are "about" or "directed toward." Mental contents are typically assumed to be propositions specifiable by "that" clauses-for example, someone might believe that "the food is located under the bucket."*

X understands, by some means, the relations between an agent's mental states, their environmental conditions, and their behavior. This understanding enables X to make predictions about others' behaviors and to explain those behaviors.

to direct the "curiosity to understand oneself" toward oneself; in the third passage, Flavell, referring to the use of the "strategy" subcategory of metacognitive knowledge, claims that the latter represents the first step in enacting the agent's manifest behavior in the world. An emblematic example might be that of a little girl who, in order to relate to her little brother, knows that she will have to adopt "strategy A" in "task X" instead of "strategy B" for "task Y", implicitly suggesting that metacognition has its own relational and intersubjective openness[44].

If this clarification convinces as to the correctness of the phenomenological explanation of metacognition, one should not be surprised if the above schematization of the various steps of metacognition, explained through the use of Flavell's thesis, is in fact the phenomenological schematization of the three key steps of the process of mentalization used by Daniel Hutto, as it is presented in the debate today.

This would thus suggest that mindreading (the latest transformation of the concept of "cognitive empathy") and metacognition are the same phenomenon, at least as far as the phenomenological aspect is concerned, and that speaking of "cognitive empathy" generates conceptual errors and misunderstandings about the way intersubjectivity between an I and a You occurs.

5. The naturalization of the I and the "relational void"

The course outlined so far has aimed to highlight the strong conceptual limitations of cognitive empathy and its possible overlap with the concept of metacognition; in order to succeed in this endeavor several key steps have been taken which, chronologically, starting from the transformation of the concept of *sympathy* in Hume have reached the meaning of "entering into the emotions of others, experiencing the same intensity"; from the objectification and abandonment of the concept of mind by *folk psychology* to its representation in the debate by Premack and Woodruff's paper; from the neurobiological ability to explain the cognitive effort of empathy through the "mindreading module" to the relationship between mentalization and metacognition, generating a further shift in the meaning of cognitive empathy.

It is difficult to find a philosophically convincing justification for the latter step: the possible overlap between mentalization and metacognition suggests that they are not two different and distant phenomena, but one and the same, interpreted from different perspectives.

Therefore, if cognitive empathy/mentalization does not seem to be a satisfactory explanation of intersubjectivity, it remains to be explained how an I can relate to a You and how it can "understand", as if it were the active protagonist of this emotional narrative, the emotions felt by the other.

An alternative explanation would be to look at the inter-relationship in a different way, starting precisely from the nature of the I or the person acting toward the other's subjectivity.

While fully sharing the program of cognitive naturalization[45] and cognitivist research in the philosophy of mind on the fragility and unreliability of the I[46], one is inclined to argue that the greatest deception to which the sense of one's own subjectivity falls victim is precisely the illusion of a simplified image of an "acting, conscious, and aware Self," constructed by the brain as the result of intense brain activity. Therefore, at the basis of intersubjectivity, there is not the (conscious, rational, in the first person) I acting toward the You, but an intense neural activity, defined as the *Self* which, to survive, gives itself an "ideal" image (the I) with which it can relate to the context. Interestingly, in recent years, the naturalized approach to the mental has been

44 Flavell, 1979.

45 Cf. Nannini, 2007.

46 Cf. Wegner, 2005, 2018; Gazzaniga, 2016.

extended to an evolutionary explanation of the mental itself: taking up "Darwin's law" on the "survival of the fittest" and on genetic mutation as the basis for the structural change in the species, the "mind phenomenon" is included within this explanatory logic; thus considered as the ultimate result of a selection process that has led from the appearance of bacteria to authors such as Johann Sebastian Bach[47]. In the wake of what has just been said, Daniel Dennett, for example, also considers the mental experience the result of an illusion arising from complex brain mechanisms that provide first-hand experience of "being a *cogito*" and "not just a body". A useful and functional stratagem, evolutionary in nature, for the survival and adaptation of a species in a given context, which would ward off dualistic explanations of the mental[48].

The first major transformation that these approaches entail in the debate on intersubjectivity is the change in the nature of the protagonists involved: no longer an "I" relating to a "You", but a *Self* relating to a You, signifying that in the above-mentioned relationship there would never be an "underlying sincerity" that guarantees that both participants understand each other's emotional states. Hence, this gives rise to the consequence on the intersubjective relationship, which completely changes its meaning, resulting in the existence of a double "void" to be filled in the *space of the relationship* between a *Self* and a You, since:

- On the *Self's* side, personal agentivity is constructed after the action of inter-relation has been implemented and, illusively, the *Self's* "consciousness" has self-narrated that it has been the author of the action "all along" (consciously, the *Self* has established a relationship with the You).
- On the You's side, in addition to the same mechanism experienced first-hand by the You itself, there is also the illusion of having communicated its desires and beliefs to the other *Self* just as the You experiences them first-hand.

This explanation is what is meant by "relational void"; the result of a structural incommunicability between interacting subjects as a result of the functional deception that the brain performs to itself as the structuring of a self-image, projected outwardly: the "I."

From this it follows that there will always be a double void to be filled, both in the acting subject and in his interlocutor; in a *Self*-You type scheme, the first "void" is created between the *Self* and the You in the space of relationship: how could the *Self* be "certain" that what it will say or do will be the original result of its intentions and not an explanation that arose only after the action has already been enacted?

The second "void" is created in the communication from the You to the *Self* of one's intentions, beliefs, and desires: how could the You be certain that what he himself wants to communicate is understood in the terms in which he himself communicates them to the *Self*?

The result of the "relational void" is the structural incommunicability between any interacting subject; an inability, therefore, to "empathize with the other," initiating a cognitive effort such as to arrive at the "reading the mind" of the other, since both the *Self* and the You are unable to "structurally step outside themselves" to understand intentions or behaviors that are carried out from a third-person perspective. In other words, neither will the *Self* be able to "come out of itself" to transfer, transport and, thus, understand its own point of view in that of the other; nor will the You ever be certain that the other has initially understood what it is communicating.

The reference to the *monad* would seem almost natural; however, the difference between Leibnitz's conception and that of the "relational void" is evident in the "solution" provided:

47 Cf. Dennett, 2017.
48 Cf. Frankish, 2023.

while for the philosopher from Leipzig the inter-relation between two or more monads took place through reference to the *pre-established harmony*, in this position intersubjectivity is possible through metacognition first; then through the shifting of one's beliefs, volitions and desires onto the other, by means of the imagination.

Therefore, one finds oneself in a state of a complete absence of relation and intersubjectivity which, however, seems in fact to be denied by the daily social relationships that everyone has: how can we explain this apparent contradiction?

The phenomenological process underlying "relational void" is able to overcome and resolve such apparent incompatibility precisely through the use of metacognition.

In fact, the *Self*, thinking that it "authentically" understands the account of the You, "narrates" itself to find itself in these same conditions; for the *Self*, the result would be to imagine the emotional sphere it would feel if it were in that hypothetical context and the relative possibility of narrating what it feels illusively. Only at this point, the use of metacognition having ceased, will the *Self* be able to "transfer" such an imaginary experience onto the You.

In this context, incommunicability derives as much from the sender, the You, as from the receiver, the *Self*, since both are unable to "come out of themselves"; the debate on cognitive empathy thus seems to have focused on the result of the complex naturalized system of encephalic activity, resorting to "mysterious" capacities or tools (empathy, cognitive empathy, mindreading) to explain intersubjectivity, emphasizing only the "immersion in feeling" and ignoring the causal process that generated it.

In the light of this explanation, the "relational void" could be outlined as follows:

1. The *Self* assumes that it has *understood* a situation of distress communicated by the You;
2. The *Self* imagines its own agentivity in the same experiential framework;
3. The *Self* understands what emotions it might experience in the same situation;
4. The *Self* transfers the hypothetical life experience onto the You, narrating to itself the hypothetical emotional sphere it might experience;
5. The *Self* relates to the You, telling it that it has "understood" the situation of distress, thinking, illusively, that it has "entered into the mind" of the You and experiencing first-hand the *same* emotional sphere as the You.

Behind such "apparent simplicity" lies a complex *mental* operation capable of providing everyone with the illusion of understanding, making oneself understood and relating to the other; once again, a "magnificent trick" that the brain "plays" on itself.

Therefore, in the light of what has been argued so far, we would like to support the abandonment of the concept of cognitive empathy and its derivatives in order to opt for a *simpler and more plausible* explanation of intersubjectivity.

Conclusions

In light of what has been discussed, this paper did not aim to conclusively demonstrate the illusory nature of cognitive empathy, but rather to raise a number of critical arguments against some of its current interpretations. The reference to the key steps in the emergence and transformation of the concept of cognitive empathy suggests that there has been a misunderstanding of the use of the term, its concept and its aims in the evolutionary history of the concept; indeed, from Hume to the present day, the very meaning of empathy has changed, from sympathy as the origin of morals, according to the Scottish philosopher, to the reading of the mind in the current debate. One possible interpretation of this change could be advanced by suggesting that the naturalization of the Cartesian *cogito* may have provided

a practical element to explain both the nature of the mental and intersubjectivity. This interpretation finds its objective confirmation in the theorization of folk psychology through the 'Jones myth' and its subsequent abandonment when the two great families of common sense psychology (the ST and the TT) encountered considerable difficulties in guaranteeing an explanation of the mind consistent with the premises from which they both proceeded.

This constant shift in meaning, and the departure from the very sense in which Hume hypothesized *sympathy*, may explain why, to this day, cognitive empathy seems to be a meaningless concept. What remains constant, however, is the explanation that one can "experience the other's emotions first-hand", as if one could actually "enter the other's mind" through mindreading in order to gain knowledge of it. It is no coincidence that, in the light of an emerging literature, attempts are being made to legitimize the process of mindreading by explaining it in terms of its relationship to metacognition.

With this in mind, an attempt has been made to show that when cognitive empathy is read through the lens of metacognition itself, in the context of the naturalization of the mental, it runs the risk of appearing as a redundant, if not misleading, concept. Indeed, by suggesting the naturalization of the mental, emphasis has been placed on the consequent illusory nature of the I, that is, the one who, it is assumed, initiates a relationship with the other, feeling "empathy" for what the You itself communicates to it. This would entail an initial structural transformation of the intersubjective relationship: it is no longer the I that relates to the You, but the *Self* that relates to the You, pointing out that the Self identifies the complex brain activity at the origin of the illusory nature of consciousness (the I) as the "unity of oneself".

The consequence of this clarification would be what has been called "relational void": the existence of a temporal void in the relationship between cerebral activity and the consequent unified image of the I, which leads to the illusion of having first consciously wished to establish a relationship with the You and, subsequently, having brought it into being. With regard to the You, in addition to the same illusion, we also add that of believing that one has "authentically" communicated the emotions of one's experience to the interlocutor; a false belief dictated by the structural incommunicability of each one, of each *monad*, to initiate any relationship, since it is not able to "come out of itself".

While the "relational void" proposed in the concluding pages is only a theoretical sketch, it is intended to present a possible alternative for rethinking the relationship between I and You, beyond the classical structures of simulation or the theory of mind.

While taking a critical stance towards some formulations of cognitive empathy, this paper does not aim to exhaust the issue or to deny the possibility that other notions – for example, those inspired by the phenomenology of perception or the 4E models – might provide a more adequate account of intersubjectivity. Rather, it aims to circumscribe and challenge a conceptually fragile use of cognitive empathy in the hope of stimulating a more analytical discussion of the theoretical assumptions that underpin its use.

REFERENCES

Almeida, R. et al. (2024). Neural Correlates of empathy: a systematic review of event-related potentials studies in perceptual tasks. *Brain Sciences*, 14(5), 504;

Aaltola, E. (2014). Affective empathy as core moral agency: psychopathy, autism and reason revisited. *Philosophical Explorations*, 17:1, pp. 76-92;

Baron-Cohen, S. (2025). *Zero Degrees of Empathy: A new theory of human cruelty and kindness*. London: Penguin;

Cameron, C. D., Conway, P., & Scheffer, J. A. (2022). Empathy regulation, prosociality, and moral judgment. *Current Opinion in Psychology*, Vol. 44, pp. 188-195;

Cappiello, I. (1983). *La morale della simpatia in David Hume*. Napoli: Liguori;

Carruthers, P. (2009). How we know our minds: The relationship between mindreading and metacognition. *Behavioral and Brian Sciences*, 32, pp. 121-138;
Cuff, B., Brown, S. J., Taylor, L. & Howat, D. (2016). Empathy: a review of the concept. *Emotion Review*. Vol. 8, Issue 2, pp. 144-153;
Decety, J. & Ickes, W. (2009). *The Social Neuroscience of Empathy*. Cambridge, MA: The MIT Press;
Decety, J. & Ickes, W. (2011). *These Things Called Empathy: Eight Related but Distinct Phenomena,* from *The Social Neuroscience of Empathy*. Cambridge, MA: The MIT Press;
Dennett, D. C. (2017). *From Bacteria to Back and Back: The Evolution of Minds,* New York, NY: W.W. Norton and Co. Inc.;
Dennett, D. & Huebner, B. (2009). Banishing 'I' and 'we' from accounts of metacognition. *Behavioral and Brian Sciences*, 32, pp. 148-149;
De Vignemont, F. & Singer, T. (2006). The empathic brain: How, When and why? *Trends and Cognitive Sciences*, Vol. 10, n. 10, pp. 435-441;
Eisenberg, N. (2018). Empathy and Sympathy. In M. Lewis and J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 677-691). New York, NY: Guilford Press;
Flavell, J. H. (1979). Metacognition and Cognitive Monitoring. A new Area of Cognitive-Developmental Inquiry. *American Psychologist*, Vol. 34, No. 10, pp. 906-911;
Frankish, K. (2023). What is illusionism?. *Klēsis Revue Philosophique*, n. 55, pp. 1-13;
Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17, pp. 535-554;
Gallagher, S. (2012), Empathy, Silulation and Narrative. *Science in Context*, Vol. 25, pp. 355-381;
Gazzaniga, M. (2016), *Who's in Charge? Free Will and the Science of the Brain*. London: Robinson;
Hart, J. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, pp. 208-216;
Heyes, C. et al. (2020), Knowing Ourselves Together: The Cultural Origins of Metacognition. *Trends in Cognitive Sciences*, Vol. 24, Issue 5, pp. 349-362;
Hoffman, M. (2010). *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge: Cambridge University Press;
Hume, D. (2023a). *An Enquiry Concerning the Principles of Morals*. Firenze: Paperback;
Hume, D. (2023b). *A Treatise of Human Nature*. Firenze: Paperback;
Hutto, D. et al. (2011). Editorial: Social Cognition: Mindreading and Alternatives. *Review of Philosophy and Psychology*, 2, pp. 309-315;
Jacob, P. (2011). The Direct-Perception Model of Empathy: A Critique. *Review of Philosophy and Psychology*, 2, no. 3, pp. 519-540;
Krebs, D. L. (1975). Empathy and Altruism. *Journal of Personality and Social Psychology*, XXXII, pp. 1134-1146;
Krueger, J. (2018), Direct social perception. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4e Cognition* (pp. 301-320). Oxford: Oxford University Press;
Lewis, D. (1972). Psychological and theoretical identification. *Australian Journal of Philosophy*, no. 50;
Lewis, M. & Haviland-Jones, J. M. (2018). *Handbook of Emotions*. New York, NY: Guilford Press;
Nannini, S. (2007). *Naturalismo Cognitivo. Per una teoria materialistica della mente*. Macerata: Quodlibet Studio;
Premack, D. & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind? *Behavioral and Brain Sciences*, no. 4, pp. 515-629;
Putnam, H. (1995). Mind, Language and Reality. *Philosophical Papers*, Vol. 2, pp. 70-84;
Rizzolatti, G. & Sinigaglia, C. (2006). *So quel che fai. Il cervello che agisce e i neuroni specchio*. Milano: Raffaello Cortina Editore;

Rouault, M. et al. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, Vol. 1, pp. 1-17;
Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Frontiers in Psychiatry*, 10: 126;
Stueber, K. R. (2010), *Rediscovering Empathy: Agency, Folk Psychology and the Human Sciences*. Cambridge, MA: The MIT Press;
Wegner, D. (2005). Who is the Controller of Controlled Processes. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The New Unconscious*. Oxford: Oxford University Press;
Wegner, D. (2018). *The Illusion of Conscious Will – New Edition*. Cambridge, MA: The MIT Press;
Wispè, L. (1986). The distinction between sympathy and empathy. *Journal of Personality and Social Psychology*, Vol. 50, no. 2, pp. 314-321;
Zahavi, D. (2001). Beyond empathy. Phenomenological Approaches to Intersubjectivity. *Journal of Consciousness Studies*, 8, no. 5-7, pp. 151-167;
Zahavi, D. (2014). Empathy and Other-Direct Intentionality. *An International Review of Philosophy*, Vol. 33, no. 1, pp. 129-142.