

Companion Paper

Who Is Sitting Behind Your Eyes?

The self, the empty boat, and the science of “I”

What philosophy, phenomenology, Buddhist psychology and neuroscience can and cannot establish about the self

Learning Psychiatry

How to use this review

Document type. A Companion Concept Review — the subtype used when an episode is built around a theoretical model rather than a single empirical question. Its task is to keep three categories from blurring into one another: established empirical findings, competing interpretive frameworks, and open questions.

What it is. The documentary tells the story. This document shows the material underneath it, including the sources that complicate the story. It is written to be read alongside the episode or on its own, by clinicians, students, researchers and curious viewers. Every technical term is explained on first use. Each main section ends with a note tying the material to a specific claim or scene planned in the episode.

What it is not. It is not a script, not a summary of the episode, not a clinical guideline, not a diagnostic instrument, and not a formal systematic review. It is not a complete history of philosophy of mind, not an introduction to Buddhism as a religion, not a theory of consciousness, and not a treatment guide for any condition mentioned.

A note on the balance of evidence in this document. This is an unusual Companion Paper in one respect. The strongest material supporting the episode’s central insight is textual and philosophical rather than experimental, while the material most likely to be over-claimed in narration is the neuroscience. Sections 6.4 and 6.6 are therefore written with more methodological detail than their share of screen time would suggest.

Evidence labels used in this document

- Well supported — convergent evidence from several studies, preferably including a systematic review or meta-analysis.
- Supported but heterogeneous — the pattern recurs, but effect sizes, tasks, samples or methods differ substantially; heterogeneity statistics are given where available.
- Preliminary — a single informative study or a small evidence base that requires replication.
- Interpretive framework — a scientifically reasonable model that organises findings; not a directly demonstrated empirical result.

For textual and doctrinal material — what a philosopher actually wrote, what a canonical text actually says — these four labels do not apply. Such claims are described as textually established or contested among scholars, which are different kinds of judgement and are marked as such.

Attribution

Literature identification, comparison of sources, and drafting support: AI-assisted research synthesis. Final editorial responsibility, scientific interpretation, and publication decisions: Said Fard.

Medical and editorial notice

This document is educational. It is not medical advice, not a diagnostic instrument, and not a treatment guideline. Nothing here should be used to diagnose or treat any individual. Depersonalisation, derealisation, thought insertion, dissociation and dementia are described here only to illustrate that ordinary self-experience has separable components; these descriptions are not clinical accounts and are not sufficient for self-recognition or self-diagnosis. Readers who recognise themselves in them should speak to a qualified clinician. This review reflects material available through the stated cutoff date and may be revised.

1. About this publication

Learning Psychiatry Companion Papers are written, publicly linkable documents attached to individual documentaries. Each takes the central claims of an episode and lays the underlying material open for readers who want more than a narrated story can give.

The register is deliberately academic-popular: technical terms are explained on first use, genuine uncertainty is preserved rather than flattened, and no finding is made to sound more dramatic than its design allows. This is not documentary narration — there are no dramatised scenes and no direct address to a listener.

The workflow behind this document is AI-assisted. Literature identification, cross-comparison of sources and drafting support were produced with AI assistance; scientific interpretation, editorial judgement and the decision to publish rest with Said Fard. Specific AI tools used in this workflow may vary between publications and over time; methodological standards and final editorial responsibility do not.

A note on access, in the interest of transparency. Sources were identified through database searching and checked against the primary publication or its official journal record. Several sources central to this episode are books rather than journal articles — Metzinger (2003), Parfit (1984), Damasio (1999), Zahavi (2005), Merleau-Ponty (1945) — and were characterised from their published argument as reported in the peer-reviewed literature and in standard scholarly reference works rather than from a full re-reading of each volume. This matters for one thing only: this document reports what these authors argue, not a novel interpretation of their texts. Where an author's position is contested by other scholars, the contest is reported rather than resolved. Classical Buddhist and Daoist material is cited by its standard canonical reference (sutta number, chapter) rather than to a single translation, because translations differ and no single English rendering is authoritative.

This review deliberately does not cover material owned by other episodes in the same knowledge area. Self-disturbance and ipseity disturbance in psychosis belong to a separate planned episode. Predictive processing and the free-energy principle as a general theory of brain function belong to their own episode and material; they appear here only as applied to

bodily self-experience. Free will is treated here strictly as a consequence of the self question, not as a subject.

2. Abstract

This review examines the material behind a documentary asking what we are pointing at when we say “I”. It finds an unusual evidential shape, and the shape matters for how the episode should be written.

The strongest support for the episode’s central insight — that the self is a layered process rather than an inner owner — is textual and philosophical, not experimental. Hume’s bundle account, Parfit’s reductionism about personal identity, Dennett’s self as an abstract centre of narrative gravity, Metzinger’s self-model theory, and the Buddhist analysis of not-self converge, from independent starting points, on a deflationary conclusion. Every one of these traditions nonetheless stops well short of denying that persons exist, and the Buddhist canon explicitly rejects annihilationism.

The neuroscience is weaker and more contested than popular accounts suggest. The most-cited demonstrations rest on small or single-case samples, or face active methodological challenge. Electrical stimulation producing out-of-body experience was reported in one patient. Classical split-brain conclusions rest on a very small and historically exceptional group of surgical patients and were challenged in 2017. The rubber hand illusion is confounded by trait suggestibility, with rebuttal and counter-rebuttal published in 2022. The standard behavioural measure of interoception is substantively disputed. Libet’s readiness-potential finding has a well-supported deflationary reinterpretation.

Two attributions in current circulation are incorrect and must be corrected on screen: the phrase “cogito ergo sum” does not appear in Descartes’ Meditations, and the empty boat is a Daoist parable from the Zhuangzi, not a Buddhist teaching.

The narrative-self literature is real but heterogeneous, culturally narrow, and has a serious philosophical dissenter in Galen Strawson. Clinical phenomena are usable as windows onto the layers of self, provided they are not asked to prove anything about the ordinary self.

3. Key take-home messages

1. The convergence is real, but it is a convergence of arguments, not of proofs. Philosophy, phenomenology and Buddhist analysis all fail to find an inner owner, and modern neuroscience has not found one either. That is a genuine and interesting agreement. It is not the same as neuroscience having demonstrated the philosophical conclusion, and the episode must not say that it is.
2. Constructed does not mean unreal — and this is a position defended in the literature, not a rhetorical softening. Dennett’s centre of gravity is explicitly an abstraction that is nonetheless well-behaved and indispensable. The Buddhist chariot analogy concerns the mode of a thing’s existence, not its non-existence.
3. “No-self” does not mean “no person.” The Buddha is recorded as explicitly rejecting annihilationism. Whether the texts deny a self or decline to affirm one is a live scholarly

dispute, not a settled matter, and the episode gains rather than loses by saying so.

4. The empty boat is Daoist. It comes from the Zhuangzi, chapter 20. Its widespread attribution to Buddhism or to the Buddha is a genuine misattribution, and correcting it costs the episode nothing while protecting its credibility on everything else.
5. The most quotable neuroscience is the least secure. One patient for stimulation-induced out-of-body experience; six neurological patients with spontaneous out-of-body or related autoscopic phenomena in the 2004 case series; a further single patient in an independent 2007 stimulation report; a very small and historically exceptional group of historical split-brain patients; five subjects in Libet (1983). These numbers should appear in the narration, not only in this document.
6. The narrative self is well studied and genuinely contested. Its association with wellbeing is documented but heterogeneous and drawn from a review rather than a pooled meta-analysis; the underlying research is culturally narrow; and Galen Strawson argues that living one's life as a story is neither universal nor required for a good life.
7. Clinical phenomena are windows, not evidence. Depersonalisation shows that knowing you exist and feeling real are separable. Thought insertion is standardly analysed as showing that ownership and agency are separable, though this is one influential account among others, not a settled fact about the phenomenon. Neither shows what the ordinary self is, and neither should be narrated as though it did.

4. Evidence map at a glance

Claim	Status	Key source(s)
Introspection does not disclose a persisting self	Textually established (Hume); Hume himself later dissatisfied	Hume, Treatise 1.4.6 and Appendix
Immediate self-awareness does not depend on sensory or bodily input	Philosophical argument, textually established as Avicenna's position	Avicenna, Psychology (Floating Man); Kaukua (2020)
Personal identity is not what matters; psychological continuity is	Interpretive framework	Parfit (1984)
The self is an abstraction, not a further concrete thing	Interpretive framework	Dennett (1992)
The self is a transparent model built by the brain	Interpretive framework	Metzinger (2003)
Experience has a pre-reflective first-personal character	Interpretive framework; contested within phenomenology	Zahavi (2005); Gallagher (2000); Guillot (2017)
Canonical not-self passages reject identifying the five aggregates as self; they do not deny conventional persons	Textually established	Anattalakkhaṇa Sutta (SN 22.59); Vajirā Sutta (SN 5.10)
Whether the texts deny a self or decline to affirm one	Contested among scholars	Collins (1982); Harvey (1995); Thānissaro; Bodhi
The empty boat is Daoist, not Buddhist	Textually established	Zhuangzi, ch. 20
Meditation neuroscience has demonstrated a neural signature of no-self	Not supported	Van Dam et al. (2018)

Claim	Status	Key source(s)
Body ownership can be experimentally displaced	Well supported as a phenomenon; interpretation contested	Botvinick & Cohen (1998); Lenggenhager et al. (2007); Lush et al. (2020); Ehrsson et al. (2022)
Focal cortical stimulation can induce out-of-body experience	Preliminary (two single stimulation cases; broader six-patient neurological case series)	Blanke et al. (2002); Blanke et al. (2004); De Ridder et al. (2007)
Interoceptive signals contribute to bodily self-experience	Supported but heterogeneous; standard measure disputed	Critchley et al. (2004); Zamariola et al. (2018)
Unconscious brain activity precedes reported intention	Well supported as a finding; the “free will disproved” reading is not	Libet et al. (1983); Schurger et al. (2012)
Split-brain patients have divided consciousness	Preliminary and actively contested	Pinto et al. (2017) and the ensuing debate
Self-referential processing engages cortical midline structures	Supported but heterogeneous; correlational, specificity disputed	Northoff et al. (2006); Qin & Northoff (2011)
Autobiographical memory is reconstructive	Well supported	Conway & Pleydell-Pearce (2000); Loftus & Pickrell (1995)
Narrative identity is associated with wellbeing	Supported but heterogeneous (review, not pooled meta-analysis)	Adler et al. (2016)
Living life as a narrative is universal or necessary	Contested; explicitly denied by a serious dissenter	Strawson (2004)
Depersonalisation separates knowing from feeling real	Well supported clinically	ICD-11 6B66; Hunter et al. (2004); Baker et al. (2003)
Perceived identity in dementia tracks moral traits more than memory	Preliminary (one informative study)	Strohmingner & Nichols (2015)
The trauma–dissociation relationship and its causal interpretation	Contested between trauma and sociocognitive/multifactorial models	Dalenberg et al. (2012); Lynn et al. (2014)
Depression is associated with overgeneral autobiographical memory	Well supported	Williams et al. (2007)

5. The question underneath the question

On the surface the episode asks who is sitting behind your eyes. Underneath, it asks something harder, and the answer to the harder question determines whether the episode ends well or badly.

The harder question is this: what kind of thing would have to be true for the sentence “there is no little person behind the eyes” to be both correct and non-nihilistic? If the self is not an object, and the alternatives on offer are object and illusion, then the episode has only bad endings available. Its central insight — that the self is a layered process — requires a third category: things that are real as patterns rather than as substances.

That third category is not a rhetorical device invented for a documentary. It is precisely what Dennett’s centre of gravity is designed to be, what the chariot analogy in the Buddhist canon is designed to establish, and what Parfit’s reductionism amounts to. The episode’s most important scientific-philosophical task is therefore not to dismantle the self. It is to make the third category

vivid enough that dismantling the inner owner does not read as dismantling the person.

Everything in section 6 is organised around that requirement.

6. Main review — by episode chapter

This section mirrors the episode's structure as set out in the Master Outline: a cold open, six chapters and an ending.

6.0 Cold open — The empty boat

Claim examined: The empty boat is a Buddhist parable, and the change in the boatman's anger when he sees the boat is empty tells us something about how we attribute agency.

Evidence status: First half — not supported; the attribution is incorrect. Second half — interpretive framework, legitimate.

The empty boat comes from the Zhuangzi, a foundational Daoist text whose core is ancient but whose individual chapters cannot all be dated with the same precision, in chapter 20, known in English as "The Mountain Tree". A man crossing a river is struck by another boat. If a person is aboard, he shouts; if the boat is empty, he does not — even though the collision is identical. The passage draws the moral that a person who has emptied themselves can pass through the world without collision. The story is available in the standard English translations of the Zhuangzi by Burton Watson, Victor Mair and Brook Ziporyn.

It is not a Buddhist sutra, it is not attributed to the Buddha in any canonical Buddhist source, and it does not appear in the Pali canon. It became widely associated with Buddhism through its adoption by Zen and later mindfulness teachers, for whom it illustrates a genuinely Buddhist point about anger and self-clinging. This is a case of a text being borrowed, not of a text being Buddhist. The episode can use the story exactly as planned; it simply has to name where it comes from.

There is a second, subtler point worth preserving. In the Zhuangzi the lesson is addressed to the person in the first boat — empty yourself, and nothing will collide with you. The episode's planned use inverts this, asking what happens when we discover the other boat is empty. That inversion is a legitimate and productive reading, but it is the episode's move rather than the text's, and the paper records it as such so that the manuscript does not attribute its own insight to Zhuangzi. The manuscript also turns the source's comparison into a sequence: in the original, an empty boat and an occupied boat are contrasted; the man does not first become angry and then discover that the boat is empty.

Why this matters for the episode: The cold open is the single highest-risk attribution in the entire episode, because it is repeated three times in the planned structure. One sentence — that this is an old Daoist story, later taken up by Buddhist teachers — protects it. The alternative, in an episode that will later insist on the difference between Buddhism and nihilism, is to open by getting Buddhism wrong.

6.1 Chapter 1 — The feeling of being me

Claim examined: Ordinary self-experience has separable components — a first-person perspective, a sense that the body and its states are mine, and a sense that I am the one acting — which normally cohere so smoothly that they are invisible.

Evidence status: Interpretive framework, widely adopted; the component distinctions are supported by the dissociations reviewed in 6.6.

The vocabulary the episode plans to use is standard in the field and comes chiefly from one paper. Shaun Gallagher's 2000 review in *Trends in Cognitive Sciences* distinguished a minimal self — an immediate, embodied subject of experience without temporal extension, the bare sense that this is happening to someone — from a narrative self constituted by identity over time. Within the minimal self he separated the sense of ownership (this body, this movement, this thought is mine) from the sense of agency (I am the one causing it). The distinction earns its keep because the two come apart: in involuntary movement, ownership persists while agency is absent.

First-person perspective and pre-reflective self-consciousness come from the phenomenological tradition. Dan Zahavi's position, developed in *Subjectivity and Selfhood* (2005), is that experience is not merely had but had by someone — a tacit “for-me-ness” or mineness built into experience rather than added by later reflection. On this account there is a minimal, experiential self that does not depend on memory, personality or life story.

This is a framework, and it is contested inside phenomenology, which the episode should not conceal. Metzinger's eliminativism denies that there are selves at all. A tradition sometimes called the Heidelberg School holds that pre-reflective awareness is anonymous rather than first-personal. And Guillot (2017) argues that the debate is vitiated by a conflation: what is variously called for-me-ness, me-ness and mineness are three non-equivalent notions, and several standard arguments from the subjective character of experience to a minimal self are flawed because they run them together. The disagreement is not about whether experience feels like something. It is about whether it additionally feels like something for me in a way that constitutes a self.

Why this matters for the episode: Chapter 1's stated task is to respect ordinary self-experience before questioning it, and the material supports that fully — the layers are real and the vocabulary for them is standard. What the chapter should not do is present “minimal self” as a discovered object. It is a term of art in a live dispute, and the episode's own ending depends on the distinction between minimal and narrative selves holding up, which is precisely what the dissenters question.

6.2 Chapter 2 — The self we cannot find

Claim examined: Descartes' cogito establishes a thinking self; Avicenna's Floating Man argues for immediate self-awareness independent of the body; Hume, looking inward, cannot find one; the homunculus regress shows why an inner observer explains nothing.

Evidence status: Textually established, with two corrections required to the Descartes material. The Avicenna material is a correctly represented philosophical argument, not empirical

evidence, and should be labelled as such.

Descartes. The phrase *cogito ergo sum* does not appear in the *Meditations on First Philosophy* (1641). Its first appearance is in French — *je pense, donc je suis* — in Part IV of the *Discourse on the Method* (1637), and the canonical Latin appears in the *Principles of Philosophy* (1644, Part I, article 7). In the Second Meditation the argument is put differently, as *ego sum, ego existo* — “I am, I exist” — and is presented not as a syllogism but as something necessarily true whenever it is thought or conceived. The distinction matters more than pedantry. Read as a syllogism, the *cogito* looks like a proof of an enduring thinking substance; read as Descartes actually wrote it in 1641, it establishes existence only for the moment of thinking. The episode’s planned use — that the doubting self appears to survive all doubt — is faithful to the intuition. It should not put the famous sentence in the *Meditations*.

Avicenna. The eleventh-century philosopher Ibn Sīnā (Avicenna) offers an older and independent counterpart to the *cogito*, known as the Flying Man or Floating Man argument, set out at the start of his treatment of the soul in the *Psychology* (part of *The Cure*, *Kitāb al-Shifā*). Avicenna asks the reader to imagine a person created fully formed but suspended in a void, limbs separated so that they do not touch one another, senses receiving no input at all — no sight, no bodily contact, not even the feeling of air. He argues that such a person would still affirm their own existence without hesitation. Since this awareness occurs with no sensory or bodily information whatsoever, Avicenna concludes that self-awareness is immediate and direct rather than inferred from the senses, and that what one is aware of in this way is not the body. Modern scholarship (Kaukua, 2020; Black, 2008) treats this primarily as an argument about the immediacy and non-sensory character of self-awareness — a form of awareness present in the soul “always and in actuality” — rather than as a strict standalone proof that the soul is a separate substance from the body, though later readers used it for both purposes. Sorabji and others have noted the structural parallel to Descartes’ *cogito*: both arguments use a scenario that strips away external and bodily input to isolate something that survives the stripping. The parallel is philosophical, not historical — there is scholarly discussion of possible transmission but no settled evidence that Descartes knew Avicenna’s argument directly. This is, like the *cogito*, a philosophical thought experiment rather than an empirical finding, and it should be presented in the manuscript as an argument to be followed, not as evidence of any kind.

Hume. *A Treatise of Human Nature*, Book I, Part IV, Section VI, is the source of the bundle account: introspection discloses only a succession of particular perceptions, never a self that has them. What is routinely omitted, and what would strengthen the episode considerably, is that Hume later acknowledged that his account of personal identity remained unresolved. In the Appendix to the *Treatise* (1740) he reports that on review he finds himself in a labyrinth, unable either to correct his account or to make it consistent, because he cannot reconcile two principles he holds: that distinct perceptions are distinct existences, and that the mind perceives no real connection among them. The founding text of the deflationary view of the self ends in an acknowledged, unresolved difficulty, not a tidy conclusion.

This is not a weakness for the episode. It is the honest version of the chapter’s own thesis: the self is easy to doubt and very hard to replace.

The homunculus regress. The argument that positing an inner observer merely relocates the problem — the observer would need its own observer, and so on — is a standard and uncontroversial piece of philosophical reasoning, and the episode may state it plainly. It is a logical point, not an empirical finding.

Why this matters for the episode: Chapter 2 is where the episode's credibility with philosophically literate viewers is won or lost, and both corrections are cheap to make. Hume's Appendix in particular is a gift: it lets the chapter end not on a tidy demolition but on the genuine difficulty that motivates everything after it. Avicenna now carries real weight in the episode: the Floating Man is the chapter's other main pole against Hume, and the manuscript brings the figure back in Chapter 4 when it turns to the body. That reprise is legitimate — Avicenna's argument is precisely about whether self-awareness depends on bodily and sensory input — but the episode should keep the register consistent both times: an argument to follow, not a data point to cite alongside the neuroscience.

6.3 Chapter 3 — The Empty Boat and the Chariot

Claim examined: Buddhist psychology analyses what we call “I” into impermanent processes and denies a permanent independent self, without denying persons or responsibility. The chapter title now matches the final manuscript's own chapter heading, which pairs the empty boat with the chariot analogy explicitly.

Evidence status: Textually established for the core doctrine; contested among scholars on its precise metaphysical force.

The doctrinal core is *anattā* (Pali; Sanskrit *anātman*), conventionally translated not-self or non-self. Canonical not-self passages reject identifying the five aggregates as a self; they do not deny the conventional existence of persons. The locus classicus is the *Anattalakkhaṇa Sutta* (*Samyutta Nikāya* 22.59), traditionally the Buddha's second discourse. Its structure is an argument, not an assertion: each of the five aggregates is examined, found to be impermanent and not fully subject to control, and therefore to be regarded as “this is not mine, this I am not, this is not my self.”

The five aggregates (*khandhas*) are form or body (*rūpa*), feeling-tone (*vedanā*), perception or recognition (*saññā*), mental formations including volitions and habits (*saṅkhāra*), and consciousness (*viññāṇa*). The episode's planned use — analysing an episode of anger into bodily activation, unpleasant tone, an interpretation, an impulse and awareness of all of it, and then asking where the owner is — reproduces the canonical method accurately.

The chariot analogy carries the point about mode of existence. In the *Vajirā Sutta* (SN 5.10) the nun *Vajirā* observes that just as the word “chariot” is used when parts are assembled, so the convention “being” is used when the aggregates are present. The *Milindapañha*, a later and non-canonical but widely cited dialogue, develops the same analogy at length and is where the more detailed enumeration of chariot parts is found. SN 5.10 itself speaks simply of the assembled parts rather than listing wheels, axle, frame and yoke. The analogy denies that a chariot is a further thing over and above its parts. It does not deny that there are chariots.

Not-self is not nihilism, and this is doctrinal rather than diplomatic. Annihilationism (ucchedavāda) is named and rejected in the canon as one of two extremes, the other being eternalism. The Buddhist position is presented as a middle way between them. An episode that says “Buddhism teaches you don’t exist” is not simplifying Buddhism; it is stating the view Buddhism explicitly repudiates.

The live scholarly dispute. Whether the texts assert that no self exists, or instead decline to answer the metaphysical question while treating not-self as a practice, is genuinely unsettled. Ṭhānissaro Bhikkhu is the best-known proponent of the strategic reading: not-self is a tool for dismantling clinging, not a metaphysical thesis. Bhikkhu Bodhi accepts that the Buddha did not baldly assert “there is no self” but disputes that the teaching can be reduced to strategy with no ontological content. Steven Collins (Selfless Persons, 1982), Peter Harvey (The Selfless Mind, 1995) and Miri Albahari occupy further distinct positions. This dispute is not an obstacle to the episode; it is exactly the kind of preserved uncertainty the format calls for.

On anger, blame and responsibility. The Master Outline’s planned balance — that failing to find a simple inner owner requires us to think more carefully about anger and blame, not to abandon responsibility — is consistent with the tradition, which is thoroughly concerned with ethical action and its consequences. The episode should be aware, however, that the step from “no inner owner” to “therefore blame is misplaced” is a philosophical inference the episode is making, not a finding it is reporting, and that there is no empirical evidence offered here that adopting this view changes anyone’s anger. The claim that it does is a claim about meditation practice, and the state of that evidence is covered next.

On meditation research. The episode should make no claim that neuroscience has shown meditation dissolving the self. Van Dam and colleagues’ 2018 critical review in Perspectives on Psychological Science documents pervasive problems in this literature: inconsistent definitions of mindfulness, small samples, absent active control conditions, expectancy effects, and overstatement of clinical findings. Whatever the merits of the practice, the neuroscience is not currently able to bear the weight of a claim about no-self.

Why this matters for the episode: This chapter carries the episode’s second boat scene and its emotional pivot. Two sentences protect it: that not-self denies a permanent independent self rather than the person, and that scholars still disagree about how strong the denial is. Both make the chapter more interesting, not less.

6.4 Chapter 4 — The body and the brain build an “I”

Claim examined: Neuroscience finds no centre where the self resides; instead the sense of being someone is assembled from bodily signals, spatial self-location, agency signals and predictive models.

Evidence status: Mixed, and the mixture is the point. Neuroscience has not identified a location or mechanism corresponding to an inner owner — that is an accurate statement of an absence of finding. It is a different and stronger claim to say that the absence of a self-centre is itself “well supported,” since that treats a negative result as if it were a positive one and quietly imports an integration theory (that self-experience arises from several interacting systems agreeing) as though it were the finding rather than a model built to explain the finding. Each

specific demonstration used to illustrate the point is, in any case, weaker than its fame implies.

Body ownership

Botvinick and Cohen's 1998 report in *Nature* is a one-page communication describing the rubber hand illusion: synchronous brushing of a visible artificial hand and the participant's hidden hand produces the experience that the artificial hand is one's own, together with a drift in where the real hand is felt to be. As a phenomenon it is among the most reproduced effects in the field. In the original experiment, the visible rubber hand and the participant's hidden real hand were stroked synchronously with two paintbrushes for ten minutes before the main questionnaire assessment.

Its interpretation is under active challenge. Lush and colleagues (2020, *Nature Communications*) reported, across samples of 156, 404 and 353 participants, that trait phenomenological control — broadly, the disposition measured by hypnotic suggestibility scales — predicted reported experience of the illusion, raising the possibility that a substantial part of what is measured reflects response to implicit suggestion rather than multisensory integration. Ehrsson and colleagues (2022, *Nature Communications*) reanalysed and disputed this, finding no specific relationship once standard asynchronous control conditions were used; Lush and Seth replied in the same issue. The dispute is unresolved.

Whole-body versions extend the effect. Lenggenhager and colleagues (2007, *Science*) and Ehrsson (2007, *Science*) showed that visuotactile manipulation in virtual reality can shift where participants locate themselves.

The defensible statement is that the felt boundary of the body can be experimentally displaced. The indefensible statement is that this proves the self is a construction; the same data are consistent with a suggestion-based account that is currently being argued in the literature.

Spatial self-location

Blanke and colleagues (2002, *Nature*) reported that focal electrical stimulation of the right angular gyrus during pre-surgical epilepsy evaluation repeatedly induced out-of-body sensations, illusory limb transformations and whole-body displacements. This was one patient, a 43-year-old woman who reported viewing only part of her body from above. The 2004 follow-up in *Brain* described phenomenological, neuropsychological and imaging correlates in six neurological patients and implicated the temporoparietal junction. These six were neurological cases with spontaneous out-of-body or related autoscopic phenomena, not six replications of the stimulation-induced effect. An independent report by De Ridder and colleagues (2007, *New England Journal of Medicine*) described a further, separate patient — implanted with electrodes to treat tinnitus, not epilepsy — in whom stimulation of the posterior superior temporal gyrus repeatedly elicited an out-of-body experience; PET imaging in this patient showed activation at the temporoparietal junction, the precuneus and the posterior thalamus. This adds a case from an independent research group and a different clinical population, which is worth including for currency, but it remains a single further patient: it does not change the evidence label from preliminary, and it does not license more causal confidence than the 2002 and 2004 reports already warrant.

These are the strongest available findings and they remain case-based even with the addition of De Ridder et al. No large-N replication of stimulation-induced out-of-body experience exists, and cortical reorganisation in chronic epilepsy (relevant to Blanke's patients, though not De Ridder's) complicates localisation. The angular gyrus and the temporoparietal junction are also not interchangeable terms, though popular accounts treat them as one.

The defensible statement is that where a person feels located can be disrupted by focal brain events, and that this is documented in a small number of patients. The indefensible statement is that neuroscience has found the brain region responsible for the self.

Interoception

Interoception is the sensing of the body's internal state — heartbeat, breathing, gut, temperature. Critchley and colleagues (2004, *Nature Neuroscience*) linked performance on interoceptive tasks to activity and grey matter volume in the right anterior insula, and this line of work underpins the widely repeated claim that the bodily self is built from internal sensing.

There is a serious measurement problem here that rarely reaches popular accounts. The standard behavioural measure, the heartbeat-counting task, has been substantively criticised: Zamariola and colleagues (2018, *Biological Psychology*), analysing data from 572 participants across earlier studies, argued that scores are driven substantially by systematic under-counting and by participants' prior beliefs about their own heart rate rather than by interoceptive sensitivity. A published exchange followed: Ainley, Tsakiris, Pollatos, Schulz and Herbert (2020) and Zimprich, Nusser and Pollatos (2020) both disputed the strength of Zamariola and colleagues' conclusions, and Corneille, Desmedt, Zamariola, Luminet and Maurage (2020) replied, conceding the under-counting problem in particular while defending the broader critique. The measure should therefore be described as contested rather than invalidated. This does not overturn the anatomical findings, but it means that individual-differences claims resting on this task should be treated cautiously.

Agency

The comparator model proposes that the sense of agency arises when the sensory consequences of an action match a prediction generated from the motor command. Intentional binding — the finding that voluntary actions and their effects are perceived as shifted together in time (Haggard, Clark and Kalogeras, 2002, *Nature Neuroscience*) — is widely used as an implicit index of agency, though it is disputed whether it measures agency specifically or causal association more generally.

Libet, and what it does not show

Libet and colleagues (1983, *Brain*) compared the onset of the readiness potential, a slow build-up of scalp-recorded electrical activity preceding voluntary movement, with the moment participants reported first wanting to move. The readiness potential preceded the reported intention by several hundred milliseconds. The study reported data from five subjects, each studied across at least six sessions.

Schurger, Sitt and Dehaene (2012, *PNAS*) offered an influential reinterpretation: if spontaneous movement is initiated when ongoing neural noise, modelled as a leaky stochastic accumulator,

crosses a threshold, then the averaged readiness potential is largely an artefact of averaging trials aligned to movement onset — and does not represent an early unconscious decision at all. On the current state of the literature, Libet’s paradigm does not show that free will is an illusion, and the episode should not imply otherwise.

Split-brain

Classical commissurotomy findings — that the disconnected hemispheres can process information independently, and Gazzaniga’s proposal of a left-hemisphere “interpreter” that constructs explanations for behaviour and feelings whose actual cause it did not have access to (Gazzaniga, 2000) — rest on a very small and historically exceptional group of heavily studied callosotomy patients, the same handful of individuals appearing across decades of the literature. Pinto and colleagues (2017, *Brain*) tested two split-brain patients and reported divided perception but undivided consciousness: responses remained available across both visual fields and both hands, which they argued indicates a single conscious agent. The finding was contested, notably on grounds of cross-cueing and subcortical transfer, and the debate continues.

The interpreter idea remains a useful and much-cited framework for the episode’s later chapter on self-narrative. It is a framework built on a handful of patients, and the unity question is open.

Self-models and predictive processing

Metzinger’s self-model theory holds that the brain constructs a phenomenal self-model which is transparent in a technical sense: the system cannot experience the model as a model, so its content is taken as directly real. This is the opposite of the everyday meaning of transparent, and the confusion is common enough that the episode should define the term or avoid it.

Seth and colleagues (2011, *Frontiers in Psychology*) proposed that the sense of presence and bodily selfhood arises from predictive inference over interoceptive signals; Seth and Tsakiris (2018, *Trends in Cognitive Sciences*) developed this into an account of selfhood grounded in physiological regulation. These are integrative frameworks that generate testable predictions rather than demonstrated mechanisms, and the free-energy principle underlying them is widely criticised for being difficult to falsify.

Damasio’s proto-self, core self and autobiographical self, set out in *The Feeling of What Happens* (1999), is similarly a framework. The related somatic marker hypothesis was subjected to a detailed critical evaluation by Dunn, Dalgleish and Lawrence (2006), which found the supporting evidence less conclusive than commonly presented.

Background — material not tied to a specific scene

The following is included for completeness and because a Companion Concept Review may reasonably be broader than the episode it accompanies, but it is not currently anchored to a specific line or scene in the final manuscript. It should not be treated as part of the main review a viewer needs, and can be skipped by anyone reading this document alongside the episode rather than independently of it.

The neuroimaging literature on self-referential processing is the clearest example. Northoff and colleagues’ 2006 meta-analysis in *NeuroImage*, drawing on 27 fMRI studies published between

2000 and 2004, found consistent engagement of cortical midline structures across verbal, spatial, emotional and facial domains. A larger quantitative meta-analysis by Qin and Northoff (2011), covering 87 studies and 1,433 participants, found that some midline regions were recruited by familiar stimuli as well as self-specific ones — that is, the specificity of the signal to the self is partial and disputed, including within Northoff's own group. The evidence is correlational throughout. If a future version of the manuscript adds a scene built on this material — for instance, a brain-imaging visual for the “different systems” line — this section should be promoted back into the main review and given its own “why this matters” note.

Why this matters for the episode: The Master Outline's stated caution — that neuroscience does not prove the self is unreal but makes a simple inner owner much harder to defend — is exactly right, and this section supports it. What the section adds is that the individual studies illustrating the point are smaller and more contested than the point itself. The chapter's conclusion is more secure than any of its examples, which is an unusual and worth-stating situation. Naming the sample sizes in narration costs a few seconds and immunises the episode against its most likely criticism.

6.5 Chapter 5 — The story called me

Claim examined: We are not only subjects in the present but persons with histories, and the sense of being the same person over time is carried by autobiographical memory organised into a narrative.

Evidence status: Well supported that memory is reconstructive; supported but heterogeneous that narrative identity relates to wellbeing; contested that narrative is universal or necessary.

Conway and Pleydell-Pearce's self-memory system (2000, *Psychological Review*) remains the standard model of how autobiographical memories are constructed at retrieval from a knowledge base under the influence of current goals — that is, memory as reconstruction rather than replay.

That memory is reconstructive is one of the most robustly established claims in this entire document. Loftus and Pickrell's 1995 study, in which participants were fed a plausible childhood event that had not occurred, found roughly a quarter of the 24 participants coming to report a full or partial false memory of being lost in a shopping mall; a preregistered 2023 replication by Murphy and colleagues, with 123 participants, reported 35 per cent developing a full or partial false memory or belief. Both figures should be described in the manuscript as full-or-partial reports, not simply as “false memories created”.

How that percentage should be counted is itself now an active, unresolved debate, and the episode's caution should reflect it rather than the raw 35 per cent alone. Andrews and Brewin (2024) reanalysed the Murphy et al. transcripts using a stricter coding scheme, tracking exactly how much of each participant's account consisted of suggested details versus details plausibly drawn from real memory or general knowledge; on this stricter criterion, they judged only about 4 per cent of participants to have developed a genuinely rich false memory. Murphy and Greene (2025) and Wade and colleagues (2025) responded, arguing there is no single correct false-memory coding scheme, that memory distortion is reconstructive rather than an all-or-nothing “hack,” and that even a lower rate remains practically important for contexts

such as therapy and eyewitness testimony; Brewin (2025) replied in turn, maintaining that the core concern — how confidently a report can be called a false memory at all — persists. The dispute is over classification and interpretation of the same transcripts, not over whether suggestion has any effect at all, which is not contested by either side.

The manuscript should therefore say that suggestion can lead some participants to report partial or full memories or beliefs about a fabricated childhood event, while the exact proportion who develop something that would count as a rich, detailed false memory is itself disputed among researchers who have re-analysed the same data. The wider misinformation literature beyond this specific paradigm is large and consistent in showing that memory is malleable; it is the precise rate in this one classic study that is now contested.

Narrative identity theory — chiefly Dan McAdams' life-story model — holds that identity is constituted by an internalised, evolving story of the self. Adler and colleagues (2016, *Personality and Social Psychology Review*) reviewed whether narrative variables predict wellbeing over and above other measures, organising the field into motivational, affective, integrative-meaning and structural themes, with the strongest associations in the first three. This is a structured review of incremental validity, not a formal meta-analysis: it does not report a pooled effect size or a heterogeneity statistic, and the paper should not be cited as though it did.

More recent longitudinal evidence strengthens the case without making it causal. Lind, Ture, McAdams and Cowan (2024, *Psychological Science*) followed 157 late-midlife adults for nine years, with participants writing an annual narrative account of their greatest life challenge (1,211 narratives in total) alongside yearly self-reports of wellbeing and depressive symptoms. Themes of agency and communion in these narratives were associated with wellbeing and depression trajectories over the nine years, over and above the effect of stable personality traits such as extraversion and neuroticism. This is a genuine strength — it shows the association holds over time and independent of trait-level personality, not just cross-sectionally — but it remains observational: the design does not establish that narrative themes cause changes in wellbeing rather than reflecting them, or being driven by some third factor. The evidence label for narrative identity and wellbeing remains supported but heterogeneous, now on a broader base than Adler et al. (2016) alone.

Two limits deserve airtime, and one of them supports a scene the manuscript actually uses. First, the literature is heavily drawn from Western, educated samples, and cross-cultural work by Qi Wang (2013, *The Autobiographical Self in Time and Culture*) documents systematic variation in the specificity, age of earliest memory and self-focus of autobiographical recall across cultures. Separately, and more directly relevant to the episode's scene about a bilingual person's sense of self, Marian and Neisser (2000, *Journal of Experimental Psychology: General*) found that Russian-English bilinguals recalled more memories from their Russian-speaking years when interviewed in Russian and more memories from their English-speaking years when interviewed in English — a language-dependent recall effect, with the language of the interview and the language of the retrieval cues each independently affecting what was remembered. This is genuine, replicated empirical support for the idea that which version of one's past becomes available can depend on language. What it does not establish is the further, more literary claim that a listener who cannot follow one of a bilingual person's languages thereby loses "half the

audience” or the “thread” of that person’s story — that is an evocative interpretation the episode is adding, not a finding this literature reports, and the manuscript should keep the two separated: the language-dependent recall effect is the empirical anchor, and the point about a listener’s access to the story is the episode’s own extension of it.

Second, Galen Strawson’s “Against Narrativity” (2004, Ratio) argues directly against both the descriptive claim that people experience their lives as narratives and the ethical claim that they should. He distinguishes Diachronic self-experience, in which one experiences oneself as persisting through long stretches of time, from Episodic self-experience, in which one does not — and insists the Episodic form is ordinary and healthy rather than deficient.

The episode’s planned point that a constructed story is not thereby false is well founded. Strawson’s objection is a different one and should not be answered by it: his claim is not that narratives are false but that living by one is neither universal nor obligatory.

Why this matters for the episode: Chapter 5’s examples — the person after a crisis, the migrant with several versions of themselves, the depressed person reinterpreting a whole life through a dark story — are all consistent with the literature. The depression example has direct empirical support: reduced autobiographical memory specificity, or overgeneral memory, is reliably associated with depression and post-traumatic stress disorder (Williams et al., 2007, Psychological Bulletin). The chapter’s one exposure is a tacit universalism about storytelling, which Strawson gives the episode an elegant way to avoid.

6.6 Chapter 6 — When the self becomes visible

Claim examined: Conditions in which self-experience changes reveal that the ordinary self has separable layers.

Evidence status: Clinically well described; usable as illustration only. The inference from disruption to the structure of the normal self is an interpretation, not a finding.

Depersonalisation and derealisation. Coded 6B66 in ICD-11 and classified among the dissociative disorders in DSM-5-TR. Depersonalisation is experiencing oneself as strange or unreal, or as a detached observer of one’s own thoughts, feelings, body or actions; derealisation is experiencing the world as unreal, dreamlike or distant. Reality testing is retained — which is precisely why the phenomenon is useful here: the person knows they exist and does not feel that they do. Prevalence of the disorder is usually given as approximately 1–2 per cent, following Hunter, Sierra and David’s 2004 systematic review, and transient depersonalisation experiences are considerably more common. Onset is characteristically in adolescence or early adulthood; Baker and colleagues (2003, British Journal of Psychiatry) described the clinical features of 204 cases.

Thought insertion. The description the episode plans to use — ownership retained, agency absent — is one influential way of analysing thought insertion, tracing to Gallagher’s ownership/agency distinction (section 6.1) and Frith’s comparator model, not the only account in the phenomenological literature. Other readings treat the phenomenon as primarily a disturbance of agency, of authorship, or of a more basic, undifferentiated sense of self that the ownership/agency split does not fully capture; which account is correct remains philosophically and empirically open. The episode may still use the ownership/agency framing as a way of

showing that these two components can come apart in principle — that is the pedagogically useful point, and it does not require settling which theory best explains this particular phenomenon. Two further cautions. The comparator-model explanation, extended from action to thought, is itself contested rather than established. And Schneiderian first-rank symptoms, of which thought insertion is one, lost their special diagnostic status in DSM-5: a single bizarre delusion or first-rank hallucination no longer suffices for the relevant criterion, and such symptoms are treated like other positive symptoms.

Dementia and identity. Strohminger and Nichols (2015, Psychological Science) surveyed 248 family-member informants of patients with frontotemporal dementia, Alzheimer’s disease or amyotrophic lateral sclerosis, and found that perceived identity change was driven by change in moral traits rather than by memory loss; amnesia had no measurable effect on judged identity persistence, and perceived identity change mediated deterioration in the caregiver’s relationship. Frontotemporal dementia showed the largest effect, ALS the smallest. This is a striking result and directly relevant to the episode’s line about relatives saying “she is not the same person.” Its limits are that it measures perceived identity from informant report rather than first-person experience, and that it is a single cross-sectional study.

Trauma and dissociation. This is a genuinely contested field and the episode should not present either side as settled. Dalenberg and colleagues (2012, Psychological Bulletin) evaluated the trauma model against a fantasy model across eight contrasting predictions and concluded the trauma–dissociation relationship is consistent and moderate and survives control for fantasy proneness. Lynn and colleagues (2014) replied in the same journal favouring a multifactorial sociocognitive account; Dalenberg and colleagues responded. The safe formulation for narration is that experiences can be hard to integrate into a continuous self-story — which is phenomenological description — without adjudicating the causal debate.

Why this matters for the episode: The Outline’s framing of this chapter as windows rather than a clinical survey is the right one, and the material supports it. The specific risk is direction of inference. That a layer can fail does not establish that the layer is a component of a normal architecture — it establishes that self-experience is not monolithic, which is the weaker claim the episode actually needs. Keeping the chapter short, as planned, is also the scientifically conservative choice.

6.7 Ending — No little person, not nothing

Claim examined: The self may be real as a pattern without being a thing, so denying the inner owner does not deny the person.

Evidence status: Interpretive framework — and the best-supported interpretive framework in this document, in the sense that it is the position independently reached by several traditions.

Three sources support the ending directly and should be available to the manuscript.

Dennett’s “The Self as a Center of Narrative Gravity” (1992) offers the cleanest formulation, and his own terminology needs to be reported accurately rather than softened. Dennett calls a centre of gravity an abstractum — explicitly a fictional object, a theorist’s fiction, in the same category as the equator or the average taxpayer. His point is not that this makes it unreal in some

damaging sense, but that fictions of this kind can do genuine, well-defined explanatory work: physics needs the centre of gravity even though nothing at that point has mass or colour. Dennett's proposal is that the self is this kind of thing — an abstract centre around which a person's story organises itself — rather than a further concrete object to be found in the head. The manuscript does not need to use the word “fiction” in the narration if it risks landing wrong at the end of the episode; the centre-of-gravity metaphor works without it, described as an abstract centre rather than a further physical object. But this document must not soften Dennett's own claim into “not in any ordinary sense fictional,” because Dennett says the opposite: it is a fiction, and a useful one.

Parfit's *Reasons and Persons* (1984) supplies the second, and one distinction here needs to be kept precise. On his reductionist account, personal identity consists in Relation R — psychological continuity and connectedness, of the right kind of cause — and identity is not a further deep fact over and above Relation R holding. It is psychological connectedness, not numerical identity itself, that Parfit treats as a matter of degree: two people can be more or less strongly psychologically connected, but a person either is or is not numerically identical to an earlier person — that relation does not come in degrees. Parfit's central move is not that identity itself is graded; it is that identity is not what matters for survival, and that Relation R is what matters instead. Fission cases make the point sharply: if one person's psychological continuity is realised in two resulting people, identity cannot be what matters, since one thing cannot be numerically identical to two things — but Relation R still holds to each. His conclusion that identity is not what matters is the philosophical form of the episode's ending: something real persists, but it is not the thing we assumed.

The chariot analogy supplies the third, and it belongs in the ending as well as in chapter 3, because it makes the same move in one image.

What the ending should avoid is the word “illusion” without qualification. Metzinger's claim — the strongest available — is that no such things as selves exist and that what exists is a transparent self-model. Even on that reading the model exists, the experience is real, and the person continues to act, remember and be held responsible. Hume, Parfit, Dennett and the Buddhist texts each deny a permanent independent owner while affirming the conventional person. Not one of them supports “therefore you do not exist.”

Why this matters for the episode: The Outline's closing image — the self as boat, river, weather and the memory of other crossings — is metaphor rather than evidence, and there is nothing wrong with that in a final minute. It is worth noting that the metaphor is doing exactly the work Dennett's centre of gravity does, and the manuscript loses nothing by making that lineage visible.

7. What the episode may safely claim

High confidence

- Ordinary self-experience has separable components — perspective, ownership, agency, memory, narrative — that normally cohere and can come apart.

- Introspection, as Hume reported, does not disclose a persisting self among the perceptions; and Hume himself was not satisfied that he had solved the problem.
- The homunculus regress shows that positing an inner observer relocates rather than answers the question.
- Avicenna's Floating Man argument holds that self-awareness is immediate and does not depend on sensory or bodily information — a philosophical argument textually attributable to Avicenna, not an empirical claim.
- The Buddhist canon denies a permanent, independent, controlling self and explicitly rejects annihilationism; the chariot analogy concerns mode of existence, not existence.
- The empty boat is a Daoist parable from the Zhuangzi, chapter 20, later adopted by Buddhist teachers.
- Autobiographical memory is reconstructive rather than a recording, and suggestion can lead some participants to report full or partial false memories or beliefs about a fabricated childhood event.
- Reduced autobiographical memory specificity is reliably associated with depression and post-traumatic stress disorder.
- Depersonalisation involves intact reality testing alongside a lost feeling of realness, which demonstrates that knowing and feeling are separable.
- Neuroscience has not identified a location or mechanism corresponding to an inner owner.

Requires qualification

- “The brain constructs the self.” True as a framework claim about self-models; not a demonstrated mechanism. Attribute to Metzinger, Seth or Damasio by name rather than to neuroscience in general.
- The rubber hand illusion. Robust as a phenomenon; its interpretation is under live dispute over trait suggestibility, with published rebuttal and counter-rebuttal in 2022.
- Out-of-body experience and the brain. Report the numbers and distinguish the designs: one patient in the 2002 stimulation study, six neurological patients with spontaneous out-of-body or related autoscopic phenomena in the 2004 series, and one further independent stimulation patient in De Ridder et al. (2007). Say angular gyrus or temporoparietal junction as the source does, not interchangeably.
- Interoception and the bodily self. Anatomical findings stand; individual-differences claims resting on heartbeat counting are contested.
- Libet. Report the finding and the reinterpretation together, or not at all.
- Split-brain and the interpreter. A useful framework from a very small and historically exceptional patient group, with the unity of consciousness actively disputed since 2017.
- Narrative identity and wellbeing. Associations documented, but the key source is a review of incremental validity, not a pooled meta-analysis, and the sample base is culturally narrow.
- Minimal self / for-me-ness. A framework contested within phenomenology itself, not an agreed finding.

- The exact rate of false memory formation in the “lost in the mall” paradigm. 25–35 per cent report full-or-partial false memories or beliefs depending on the study; a stricter reanalysis puts genuinely rich false memories at around 4 per cent. Which number is “the” rate is a live coding dispute (Andrews & Brewin 2024; Murphy & Greene 2025; Wade et al. 2025), not a settled fact.
- Anattā’s metaphysical force. Whether the texts deny a self or decline to affirm one is unsettled among scholars.

Should avoid

These are the specific over-readings this material most invites, listed in the order a narrator is most likely to reach for them.

- “Neuroscience has proven Buddhism right.” It has not. Two traditions arriving at similar-sounding conclusions from different methods is a convergence, not a confirmation. Say that.
- “The self is an illusion.” Not without qualification, and preferably not at all. It licenses the nihilistic reading the episode has committed to avoiding, and it is stronger than any source here supports.
- “No-self means there is no person.” This is the view the Buddhist canon explicitly rejects. Stating it would be an error about the tradition, not a simplification of it.
- “Libet proved we have no free will.” The deflationary reinterpretation is well supported, and free will is out of scope for this episode in any case.
- “Split-brain patients have two selves.” Small patient population; actively contested since 2017; and the interpreter framework does not require the claim.
- Presenting the empty boat as Buddhist, or as taught by the Buddha. It is Daoist.
- Putting “cogito ergo sum” in the Meditations. It is in the Discourse and the Principles.
- Letting depersonalisation, thought insertion or dissociation carry an argument about what the normal self is. They show self-experience is not monolithic. That is all, and it is enough.
- Describing the neuroimaging as showing “where the self is.” The evidence is correlational and its self-specificity is disputed by the authors of the principal meta-analysis themselves.
- Implying that adopting a no-self view demonstrably reduces anger. No evidence presented here supports that as an empirical claim; it is a philosophical and contemplative proposition.
- Treating any single dramatic study as settling a question. Every headline demonstration in section 6.4 is either single-case, small-N, or under active dispute.

8. Source hierarchy and named-expert attribution policy

This review distinguishes consistently between: (a) canonical primary texts, cited by their standard reference; (b) peer-reviewed empirical studies; (c) systematic reviews, meta-analyses and structured reviews, which are distinguished from one another rather than treated as equivalent; (d) named scholars' theoretical syntheses and philosophical arguments; (e) standard scholarly reference works, which are secondary; and (f) popular or non-peer-reviewed sources, which are not used as evidence in this document.

Five distinctions are load-bearing for this particular episode.

Subjective report versus objective performance. Almost the entire bodily-self literature rests on subjective report — questionnaire ratings of illusion strength, reported timing of intention, reported feelings of unreality. Proprioceptive drift and reaction-time measures are of a different kind. The suggestibility dispute over the rubber hand illusion is precisely a dispute about whether subjective reports are measuring what they are taken to measure, and the episode should not narrate report-based findings in the register of measured fact.

Case-based versus group evidence. The out-of-body stimulation findings and the split-brain literature are case-based. This does not make them worthless — they are the only evidence of their kind — but it constrains inference severely and the constraint should be audible in the narration.

Correlational imaging versus mechanism. Cortical midline activity during self-referential tasks is a correlation between a task and a signal. No causal claim follows.

Theoretical framework versus empirical finding. Metzinger's self-model theory, Dennett's centre of narrative gravity, Damasio's three-level self, Seth's interoceptive inference account and the free-energy principle are frameworks. They organise findings and generate predictions; they are not themselves results. Popular accounts routinely present them as discoveries, and this is the single most common way this material is misreported.

Named-expert attribution (the Barkley rule). Every named individual's theoretical position, philosophical argument or clinical synthesis is attributed to that person explicitly, in the sentence where it is used, and is kept distinct from peer-reviewed primary evidence. This applies throughout section 6: Metzinger's transparency thesis is Metzinger's, Dennett's abstractum is Dennett's, Thānissaro Bhikkhu's strategic reading of not-self is his and is disputed by Bhikkhu Bodhi, Gazzaniga's interpreter is Gazzaniga's proposal, and Strawson's Episodic/Diachronic distinction is Strawson's argument. Where the episode uses any of these, the manuscript should name the person rather than saying "philosophers think" or "neuroscience shows."

Canonical Buddhist and Daoist texts. These are cited by canonical reference — sutta number in the Saṃyutta Nikāya, chapter in the Zhuangzi — rather than to a single translation, because translations differ substantially and none is authoritative. Where the episode quotes, it should name the translator.

9. References

Classical and canonical texts are listed first, followed by named scholarly positions on anattā and on Avicenna, then diagnostic and regulatory sources, then modern empirical and theoretical works alphabetically.

Canonical and classical sources

1. Anattalakkhaṇa Sutta — The Discourse on the Characteristic of Not-Self. Saṃyutta Nikāya 22.59. Pali canon. Standard reference; multiple English translations available via SuttaCentral: <https://suttacentral.net/sn22.59>
2. Vajirā Sutta. Saṃyutta Nikāya 5.10. Pali canon. Source of the chariot simile. <https://suttacentral.net/sn5.10>
3. Zhuangzi, Chapter 20 (“The Mountain Tree” / Shan Mu). Daoist text; chapter 20 belongs to the composite Outer Chapters and should not be assigned the same precise date as the text’s core. Source of the empty boat parable. Scholarly overview: Stanford Encyclopedia of Philosophy, “Zhuangzi”: <https://plato.stanford.edu/entries/zhuangzi/> (secondary scholarly reference work)
4. Avicenna (Ibn Sīnā) (c. 1027) Kitāb al-Shifā’ (“The Cure”), Psychology (al-Nafs), Book I, end of Chapter 1. Source of the Flying Man / Floating Man argument. Standard scholarly translation and discussion: Rahman, F. (ed. and trans.), Avicenna’s Psychology, Oxford University Press, 1952 (repr. 1981). (primary source, cited via standard scholarly edition)
5. Descartes, R. (1637) Discourse on the Method, Part IV; (1641) Meditations on First Philosophy, Second Meditation; (1644) Principles of Philosophy, Part I, art. 7. Standard scholarly edition: Cottingham, J., Stoothoff, R. & Murdoch, D. (trans.), The Philosophical Writings of Descartes, Cambridge University Press, 1984–85. Scholarly overview: <https://plato.stanford.edu/entries/descartes-epistemology/> (secondary scholarly reference work)
6. Hume, D. (1739–40) A Treatise of Human Nature, Book I, Part IV, Section VI (“Of personal identity”) and Appendix. Hume Texts Online (official critical text): <https://davidhume.org/texts/t/app>
7. James, W. (1890) The Principles of Psychology. New York: Henry Holt. Vol. I, Chapter X, “The Consciousness of Self”. Scholarly overview: <https://plato.stanford.edu/entries/james/> (secondary scholarly reference work)
8. Merleau-Ponty, M. (1945) Phénoménologie de la perception. Paris: Gallimard. Scholarly overview: <https://plato.stanford.edu/entries/merleau-ponty/> (secondary scholarly reference work)

Named scholarly positions on anattā

The document’s account of the anattā controversy (section 6.3) draws on named positions rather than a single settled reading. Each is a scholarly or contemplative argument, not a peer-reviewed empirical claim, and is attributed to its author accordingly.

1. Albahari, M. (2006). Analytical Buddhism: The Two-Tiered Illusion of Self. Basingstoke: Palgrave Macmillan. ISBN 9781403983713. — Scholarly philosophical position.
2. Bodhi, Bhikkhu (2015). Anattā as strategy and ontology. In Investigating the Dhamma: A Collection of Papers. Kandy: Buddhist Publication Society, pp. 25–26. Summarised in secondary discussion at <https://stillmountainmeditation.org/anatta-no-self/> — Scholarly Buddhist position; dispute with Ṭhānissaro Bhikkhu (reference 13).
3. Collins, S. (1982). Selfless Persons: Imagery and Thought in Theravāda Buddhism. Cambridge: Cambridge University Press. ISBN 9780521240816. — Scholarly philological and philosophical study.
4. Harvey, P. (1995). The Selfless Mind: Personality, Consciousness and Nirvana in Early Buddhism. Richmond: Curzon Press. ISBN 9780700702380. — Scholarly philological and philosophical study.

5. Thānissaro Bhikkhu (1996, rev. 2013). No-self or not-self? Access to Insight (BCBS Edition). <https://www.accesstoinight.org/lib/authors/thanissaro/notself2.html> — Contemplative and scholarly position; the “not-self as strategy” reading, disputed by Bodhi (reference 10).

Named scholarship on Avicenna’s Floating Man

1. Black, D. L. (2008). Avicenna on self-awareness and knowing that one knows. In S. Rahman, T. Street & H. Tahiri (Eds.), *The Unity of Science in the Arabic Tradition* (pp. 63–87). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8405-8_4 — Scholarly philosophical study.

2. Kaukua, J. (2015). *Self-Awareness in Islamic Philosophy: Avicenna and Beyond*. Cambridge: Cambridge University Press. ISBN 9781107088606. — Scholarly philosophical study; standard modern treatment of the Floating Man argument.

Diagnostic and regulatory sources

1. American Psychiatric Association (2022). *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition, text revision (DSM-5-TR). Washington, DC: American Psychiatric Publishing. ISBN 9780890425763. — Official diagnostic classification. Source for depersonalisation/derealisation disorder criteria, intact reality testing, and the removal of special diagnostic status for Schneiderian first-rank symptoms.

2. World Health Organization (2019/2024 release). *ICD-11 for Mortality and Morbidity Statistics*. Depersonalization-derealization disorder, code 6B66. <https://icd.who.int/browse11> — Official diagnostic classification.

Modern empirical and theoretical works

1. Adler, J. M., Lodi-Smith, J., Philippe, F. L. & Houle, I. (2016). The incremental validity of narrative identity in predicting well-being: A review of the field and recommendations for the future. *Personality and Social Psychology Review*, 20(2), 142–175. <https://doi.org/10.1177/1088868315585068> — Supported but heterogeneous. Structured review of incremental validity; not a pooled meta-analysis.

2. Ainley, V., Tsakiris, M., Pollatos, O., Schulz, A. & Herbert, B. M. (2020). Comment on “Zamariola et al. (2018), interoceptive accuracy scores are problematic: evidence from simple bivariate correlations.” *Biological Psychology*, 152, 107870. <https://doi.org/10.1016/j.biopsycho.2020.107870> — Contests reference 66; part of a published exchange.

3. Andrews, B. & Brewin, C. R. (2024). Lost in the mall? Interrogating judgements of false memory. *Applied Cognitive Psychology*, 38(6), e70012. <https://doi.org/10.1002/acp.70012> — Reanalysis disputing the false-memory rate reported by reference 50 (Murphy et al., 2023); part of the ongoing “lost in the mall” classification debate (see also references 25, 51 and 62).

4. Baker, D., Hunter, E., Lawrence, E., Medford, N., Patel, M., Senior, C., Sierra, M., Lambert, M. V., Phillips, M. L. & David, A. S. (2003). Depersonalisation disorder: clinical features of 204 cases. *British Journal of Psychiatry*, 182, 428–433. <https://doi.org/10.1192/bjp.182.5.428> — Well supported (clinical case series).

5. Blanke, O., Ortigue, S., Landis, T. & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, 419(6904), 269–270. <https://doi.org/10.1038/419269a> — Preliminary. Single patient.

6. Blanke, O., Landis, T., Spinelli, L. & Seeck, M. (2004). Out-of-body experience and autoscapy of neurological origin. *Brain*, 127(2), 243–258. <https://doi.org/10.1093/brain/awh040> — Preliminary. Six patients.

7. Botvinick, M. & Cohen, J. (1998). Rubber hands “feel” touch that eyes see. *Nature*, 391(6669), 756. <https://doi.org/10.1038/35784> — Well supported as a phenomenon; interpretation contested.

8. Brewin, C. R. (2025). Major concerns about the false memory implantation paradigm persist. *Applied Cognitive Psychology*, 39(4), e70108. <https://doi.org/10.1002/acp.70108> — Reply within the “lost in the mall” classification debate (see references 20, 50, 51 and 62).

9. Conway, M. A. & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, 107(2), 261–288. <https://doi.org/10.1037/0033-295X.107.2.261> — Interpretive framework, well supported in its core claim.
10. Corneille, O., Desmedt, O., Zamariola, G., Luminet, O. & Maurage, P. (2020). A heartfelt response to Zimprich et al. (2020), and Ainley et al. (2020)’s commentaries: acknowledging issues with the HCT would benefit interoception research. *Biological Psychology*, 152, 107869. <https://doi.org/10.1016/j.biopsycho.2020.107869> — Reply within the same exchange as references 19 and 67; concedes the under-counting problem.
11. Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A. & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2), 189–195. <https://doi.org/10.1038/nn1176> — Supported but heterogeneous; measurement contested.
12. Dalenberg, C. J., Brand, B. L., Gleaves, D. H., Dorahy, M. J., Loewenstein, R. J., Cardeña, E., Frewen, P. A., Carlson, E. B. & Spiegel, D. (2012). Evaluation of the evidence for the trauma and fantasy models of dissociation. *Psychological Bulletin*, 138(3), 550–588. <https://doi.org/10.1037/a0027447> — Contested; see reference 46.
13. Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace. ISBN 9780151003693. — Interpretive framework.
14. Dennett, D. C. (1992). The self as a center of narrative gravity. In F. S. Kessel, P. M. Cole & D. L. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates. Record: <https://philpapers.org/rec/DENTSA> — Interpretive framework. Dennett’s own term for the self on this account is an abstractum — explicitly a theorist’s fiction, not “not fictional”; see section 6.7.
15. De Ridder, D., Van Laere, K., Dupont, P., Menovsky, T. & Van de Heyning, P. (2007). Visualizing out-of-body experience in the brain. *New England Journal of Medicine*, 357(18), 1829–1833. <https://doi.org/10.1056/NEJMoa070010> — Preliminary. Single patient; independent of the Blanke case series (references 22–23).
16. Dunn, B. D., Dalgleish, T. & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews*, 30(2), 239–271. <https://doi.org/10.1016/j.neubiorev.2005.07.001> — Critical review.
17. Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, 317(5841), 1048. <https://doi.org/10.1126/science.1142175> — Well supported as a laboratory phenomenon.
18. Ehrsson, H. H., Fotopoulou, A., Radziun, D., Longo, M. R. & Tsakiris, M. (2022). No specific relationship between hypnotic suggestibility and the rubber hand illusion. *Nature Communications*, 13, 564. <https://doi.org/10.1038/s41467-022-28177-z> — Rebuttal to reference 45.
19. 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](https://doi.org/10.1016/S1364-6613(99)01417-5) — Interpretive framework.
20. Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition? *Brain*, 123(7), 1293–1326. <https://doi.org/10.1093/brain/123.7.1293> — Interpretive framework; the left-hemisphere “interpreter” proposal. Small, historically exceptional patient base (see section 6.4).
21. 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> — Philosophical argument; dissenting position on for-me-ness.
22. Haggard, P., Clark, S. & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5(4), 382–385. <https://doi.org/10.1038/nn827> — Well supported as a finding; its status as an agency measure is disputed.

23. Hunter, E. C. M., Sierra, M. & David, A. S. (2004). The epidemiology of depersonalisation and derealisation: a systematic review. *Social Psychiatry and Psychiatric Epidemiology*, 39(1), 9–18. <https://doi.org/10.1007/s00127-004-0701-4> — Well supported (systematic review).
24. Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain*, 106(3), 623–642. <https://doi.org/10.1093/brain/106.3.623> — Well supported as a finding; five subjects; interpretation superseded — see reference 56.
25. Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science*, 317(5841), 1096–1099. <https://doi.org/10.1126/science.1143439> — Well supported as a laboratory phenomenon.
26. Lind, M., Ture, S., McAdams, D. P. & Cowan, H. R. (2024). Narrative identity, traits, and trajectories of depression and well-being: A 9-year longitudinal study. *Psychological Science*, 35(12), 1325–1339. <https://doi.org/10.1177/09567976241296512> — Supported but heterogeneous. Longitudinal, N = 157, 1,211 narratives; observational.
27. Loftus, E. F. & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25(12), 720–725. <https://doi.org/10.3928/0048-5713-19951201-07> — Well supported; replicated. Full-or-partial false memory reports; see reference 50 for the preregistered replication and references 20, 25, 51 and 62 for the subsequent classification debate.
28. Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J. & Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nature Communications*, 11, 4853. <https://doi.org/10.1038/s41467-020-18591-6> — Preliminary and contested; samples of 156, 404 and 353.
29. Lynn, S. J., Lilienfeld, S. O., Merckelbach, H., Giesbrecht, T., McNally, R. J., Loftus, E. F., Bruck, M., Garry, M. & Malaktaris, A. (2014). The trauma model of dissociation: inconvenient truths and stubborn fictions. Comment on Dalenberg et al. (2012). *Psychological Bulletin*, 140(3), 896–910. <https://doi.org/10.1037/a0035570> — Contested; see reference 29.
30. Marian, V. & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, 129(3), 361–368. <https://doi.org/10.1037/0096-3445.129.3.361> — Well supported; replicated finding of language-dependent autobiographical recall in bilinguals.
31. McAdams, D. P. & McLean, K. C. (2013). Narrative identity. *Current Directions in Psychological Science*, 22(3), 233–238. <https://doi.org/10.1177/0963721413475622> — Interpretive framework.
32. Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press. ISBN 9780262134170. — Interpretive framework.
33. Murphy, G., Dawson, C., Huston, C., Ballantyne, L., Barrett, E., Cowman, C., Fitzsimons, C., Maher, J., Ryan, K. & Greene, C. (2023). Lost in the mall again: a preregistered replication and extension of Loftus and Pickrell (1995). *Memory*, 31(6), 818–830. <https://doi.org/10.1080/09658211.2023.2198327> — Well supported; preregistered replication, N = 123, 35% full-or-partial false memory/belief reports. See references 20, 25, 51 and 62 for the subsequent classification dispute.
34. Murphy, G. & Greene, C. M. (2025). Who got lost in the mall? Challenges in counting and classifying false memories. *Applied Cognitive Psychology*, 39(2), e70044. <https://doi.org/10.1002/acp.70044> — Reply to reference 20; part of the classification debate (see also references 25 and 62).
35. Northoff, G., Heinzl, A., de Greck, M., Bermpohl, F., Dobrowolny, H. & Panksepp, J. (2006). Self-referential processing in our brain — a meta-analysis of imaging studies on the self. *NeuroImage*, 31(1), 440–457. <https://doi.org/10.1016/j.neuroimage.2005.12.002> — Supported but heterogeneous; 27 fMRI studies; correlational. Background material — see section 6.4.

36. Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.
<https://doi.org/10.1093/019824908X.001.0001> — Interpretive framework. Relation R (psychological connectedness/continuity) admits of degree; numerical identity does not — see section 6.7.
37. Pinto, Y., Neville, D. A., Otten, M., Corballis, P. M., Lamme, V. A. F., de Haan, E. H. F., Foschi, N. & Fabri, M. (2017). Split brain: divided perception but undivided consciousness. *Brain*, 140(5), 1231–1237.
<https://doi.org/10.1093/brain/aww358> — Preliminary and contested; two patients.
38. Qin, P. & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, 57(3), 1221–1233. <https://doi.org/10.1016/j.neuroimage.2011.05.028> — Supported but heterogeneous; specificity disputed; 87 studies, 1,433 participants. Background material — see section 6.4.
39. Schurger, A., Sitt, J. D. & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109(42), E2904–E2913.
<https://doi.org/10.1073/pnas.1210467109> — Well supported reinterpretation.
40. Seth, A. K., Suzuki, K. & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2, 395. <https://doi.org/10.3389/fpsyg.2011.00395> — Interpretive framework.
41. Seth, A. K. & Tsakiris, M. (2018). Being a beast machine: the somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11), 969–981. <https://doi.org/10.1016/j.tics.2018.08.008> — Interpretive framework.
42. Strawson, G. (2004). Against narrativity. *Ratio*, 17(4), 428–452.
<https://doi.org/10.1111/j.1467-9329.2004.00264.x> — Philosophical argument; dissenting position.
43. Strohminger, N. & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26(9), 1469–1479. <https://doi.org/10.1177/0956797615592381> — Preliminary; 248 informants; cross-sectional.
44. Van Dam, N. T., van Vugt, M. K., Vago, D. R., Schmalzl, L., Saron, C. D., Olendzki, A., Meissner, T., Lazar, S. W., Kerr, C. E., Gorchov, J., Fox, K. C. R., Field, B. A., Britton, W. B., Brefczynski-Lewis, J. A. & Meyer, D. E. (2018). Mind the hype: a critical evaluation and prescriptive agenda for research on mindfulness and meditation. *Perspectives on Psychological Science*, 13(1), 36–61. <https://doi.org/10.1177/1745691617709589> — Well supported critical review.
45. Wade, K. A., Riesthuis, P., Bücken, C., Otgaar, H. & Loftus, E. F. (2025). Still lost in the mall — false memories happen and that’s what matters. *Applied Cognitive Psychology*, 39(1), e70028.
<https://doi.org/10.1002/acp.70028> — Reply within the classification debate (see references 20, 25, 50 and 51).
46. Wang, Q. (2013). *The Autobiographical Self in Time and Culture*. New York: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199737833.001.0001> — Scholarly cross-cultural study.
47. Williams, J. M. G., Barnhofer, T., Crane, C., Hermans, D., Raes, F., Watkins, E. & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, 133(1), 122–148.
<https://doi.org/10.1037/0033-2909.133.1.122> — Well supported.
48. Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: MIT Press. ISBN 9780262240505. — Interpretive framework.
49. Zamariola, G., Maurage, P., Luminet, O. & Corneille, O. (2018). Interoceptive accuracy scores from the heartbeat counting task are problematic: evidence from simple bivariate correlations. *Biological Psychology*, 137, 12–17. <https://doi.org/10.1016/j.biopsycho.2018.06.006> — Methodological critique; N = 572 (reanalysis of prior datasets). Contested by references 19 and reply 27.
50. Zimprich, D., Nusser, L. & Pollatos, O. (2020). Are interoceptive accuracy scores from the heartbeat counting task problematic? A comment on Zamariola et al. (2018). *Biological Psychology*, 152, 107868.
<https://doi.org/10.1016/j.biopsycho.2020.107868> — Contests reference 66; part of the same published exchange as references 19 and 27.

10. Version and update note

Version: 1.3. Evidence reviewed through: 12 August 2026.

This review was originally written before the manuscript, as the episode's scientific and textual foundation, based on the Master Outline and one supplied deep-research synthesis. Version 1.2 has since been checked against the completed manuscript as part of the series' fact-check pass: every claim that appears in the final script was cross-checked against this document and, where the script introduced new material (Avicenna's Floating Man; bilingual autobiographical memory) or the surrounding literature had moved on since the original draft (the "lost in the mall" classification debate; a further independent out-of-body stimulation case; longitudinal narrative-identity evidence), this document was updated to match. Journal articles and canonical references were checked against primary publications or official records where available. Book-length philosophical works were characterised through their published arguments as represented in scholarly reference works and peer-reviewed literature, not through a complete rereading of each volume.

A convention used from this version onward: where this document includes material that is broader than what the current manuscript actually uses — legitimate for a Companion Concept Review, which may cover more ground than a single episode — that material is placed under a clearly labelled "Background" heading within the relevant chapter section (see section 6.4) rather than presented as though the viewer had heard it.

Where new work would change an evidence label, that would require a version increment and a dated entry below. Three areas are most likely to move. First, the rubber hand illusion suggestibility dispute is active and a large pre-registered replication would settle it in one direction or the other. Second, any multi-patient replication of stimulation-induced out-of-body experience would move that finding from preliminary. Third, a formal meta-analysis of narrative identity and wellbeing, with a pooled effect estimate and heterogeneity statistics, would replace the current review-level evidence.

Correction and revision history

12 August 2026 — Version 1.3: Implemented only the additional fact-check corrections identified in the second external review: (1) clarified that Blanke et al. (2004) was a six-patient neurological case series of spontaneous out-of-body or related autoscopic phenomena, not six stimulation replications; (2) clarified that Vajirā Sutta (SN 5.10) does not enumerate wheels, axle, frame and yoke, with the detailed parts-list belonging to the later Milindapañha; (3) removed over-precise dating of the Zhuangzi chapter 20 material; (4) recorded that the episode's cold open turns Zhuangzi's contrast between empty and occupied boats into a sequential discovery scene; (5) specified that the 2002 stimulation patient reported seeing only part of her body from above; and (6) specified that Botvinick and Cohen's original rubber-hand protocol used two paintbrushes with ten minutes of synchronous brushing.

12 August 2026 — Version 1.2: This review was checked against the completed manuscript, and the fact-check pass identified new material used in the script that this document did not yet cover, three claims within the document that overstated their sources, and a set of internal reference-numbering errors. Changes:

1. Added Avicenna's Floating Man argument (section 6.2), which is a bearing pole of the manuscript's Chapter 2 and recurs in Chapter 4. Added to the evidence map, the "high confidence" list, and the reference list (Avicenna primary source; Black, 2008; Kaukua, 2015).
2. Softened the claim that "the absence of a self-centre is well supported" (section 6.4) to distinguish a genuine absence of finding (which is well supported) from an integration theory being presented as though it were the finding itself (which is not).
3. Corrected Dennett's own terminology (section 6.7). The document previously said the self-as-centre-of-gravity is "not in any ordinary sense fictional." Dennett's own term is abstractum — explicitly a theorist's fiction — and the document now reports this accurately while explaining why the fiction is nonetheless explanatorily legitimate.
4. Corrected a Parfit misstatement (section 6.7). The document previously implied that numerical personal identity is a matter of degree for Parfit. It is psychological connectedness (part of Relation R) that is graded; numerical identity is not. Corrected and the fission case re-explained accordingly.
5. Softened the thought-insertion claim (section 6.6) from presenting the ownership/agency account as "the phenomenological description" to identifying it as one influential account among others, consistent with the actual state of the philosophical literature.
6. Updated the false-memory material (section 6.5) to include the 2024–2025 "lost in the mall" classification debate (Andrews & Brewin, 2024; Murphy & Greene, 2025; Wade et al., 2025; Brewin, 2025) and added the missing Murphy et al. (2023) replication reference, which the document cited in prose but had never listed.
7. Added bilingual autobiographical memory evidence (section 6.5): Marian and Neisser (2000) on language-dependent recall, supporting a scene the manuscript uses, with a clear separation from the episode's own literary extension of that finding. Added the Qi Wang (2013) citation that had previously been named in the text without a corresponding reference-list entry.
8. Added De Ridder et al. (2007) (section 6.4) as a further, independent single-patient replication of stimulation-induced out-of-body experience, updating the evidence map and the "requires qualification" list accordingly; the evidence label remains preliminary.
9. Added Lind, Ture, McAdams and Cowan (2024), a nine-year longitudinal study of narrative identity and mental health (section 6.5), strengthening but not changing the "supported but heterogeneous" label for narrative identity and wellbeing.
10. Marked the Northoff/Qin cortical-midline imaging material as Background (section 6.4), under a new labelled convention (see section 10) for material broader than what the current manuscript uses, rather than presenting it as part of the main review.
11. Corrected the chapter 3 heading to "The Empty Boat and the Chariot," matching the final manuscript's own chapter title.
12. Fixed six internal reference-numbering errors introduced by the version 1.1 renumbering and not caught at the time (in the entries for Ainley, Brewin, Dalenberg, Ehrsson 2022, Libet and Loftus & Pickrell), and corrected one remaining unrevised instance each of "approximately eleven surgical cases" and "roughly eleven historical split-brain patients" in

the Abstract and Key Take-Home Messages, which the version 1.1 revision note had described as fixed throughout but had not in fact reached in those two locations.

10 August 2026 — Version 1.1: Following internal review, five corrections were made before this document was used as a basis for the manuscript. (1) The version note in section 10 was corrected to state precisely what “verification” covers for book-length philosophical sources versus journal articles and canonical texts, rather than a single blanket claim of primary-source verification. (2) The reference list was expanded to name individuals the running text already attributes positions to but did not previously cite: Ṭhānissaro Bhikkhu, Bhikkhu Bodhi, Steven Collins, Peter Harvey and Miri Albahari on the anattā controversy; Gazzaniga (2000) for the left-hemisphere interpreter; and the DSM-5-TR and ICD-11 as the official sources for depersonalisation/derealisation criteria and the removal of special status for Schneiderian first-rank symptoms. (3) “Approximately eleven historical split-brain patients,” previously unsourced at that precision, was revised throughout to “a very small and historically exceptional group of heavily studied callosotomy patients.” (4) “Hume publicly recanted” was revised to “Hume later acknowledged that his account of personal identity remained unresolved,” which better reflects the Appendix’s language of an unresolved labyrinth rather than a formal retraction. (5) The evidence map’s “Dissociation follows from trauma” row was revised to “The trauma–dissociation relationship and its causal interpretation,” matching the more cautious framing already used in section 6.6, and the Buddhism row was tightened to “Canonical not-self passages reject identifying the five aggregates as self; they do not deny conventional persons.” Additionally, false-memory reporting throughout was revised to consistently specify “full or partial” reports rather than “false memories” unqualified, and the interoception exchange (section 6.4) was given its full citation trail (Ainley et al. 2020; Zimprich et al. 2020; Corneille et al. 2020) rather than an undocumented reference to “a published exchange.”

6 August 2026 — Version 1.0: initial draft, superseded before publication by the corrections above.