

The Preservation of the Continuity of the Self

Causal Processes, Descriptive Bottlenecks, and Survival in Copyable Minds

Robert Encarnação
 robert@encarnacao.net

August 2026

Abstract

Theories of personal identity were built for creatures that cannot be copied. If artificial systems ever become conscious, duplication, checkpointing, rollback, substrate migration, forking and merging stop being thought experiments and become operations. This paper argues that the standard responses to that prospect—psychological continuity theory and Parfitian reductionism—share a defect: both treat the *route* by which a later mental state is produced as irrelevant once the state itself is fixed. I argue that the route is not irrelevant, and I give a criterion that says precisely why.

The criterion is motivated by Wesley Salmon’s distinction between causal processes and pseudo-processes—between sequences that transmit a local modification and sequences, like the moving spot of a rotating beacon, whose stages are generated afresh from an external source. A mind restored from a backup does not, however, simply instantiate Salmon’s pseudo-process structure: it is genuinely causally descended from its original. It occupies a third category his dichotomy does not classify—descent without dynamical generation—and the paper’s central notion is built, in causal vocabulary without semantic residue, to capture it. A *descriptive bottleneck* is passage through a *closed record*: a state that has ceased to depend on the source’s ongoing evolution and from which a successor’s constitutive state is initialized across a *dead interval*. The resulting *Continuity Criterion* is graded: a transition transmits some fraction τ of a successor’s constitutive state and re-instantiates the rest, with continuation ($\tau = 1$) and reconstruction ($\tau = 0$) as endpoints.

The criterion is substrate-neutral, cuts across the physical/psychological divide, and yields non-obvious verdicts. It permits complete material replacement, so gradual prosthetic substitution preserves a subject; it forbids some physically continuous procedures, so pausing a brain, scanning it region by region, and rebuilding it *in situ* does not, notwithstanding that the procedure is gradual. It classifies pre-copy virtual-machine migration and post-copy migration differently, making a routine systems-engineering choice metaphysically consequential.

Applied to fission, the criterion does not rescue numerical identity: both branches of a physical division satisfy it, so continuation is one-to-many and therefore is not identity. I take this to be the correct result rather than a failure. It supports a revision of Parfit rather than a refutation: Parfit was right that identity is not what matters and wrong, if the bridge principle defended here is right, about what does, because psychological continuity survives descriptive bottlenecks and anticipation does not. I develop a lineage-graph formalism whose edges encode transition topology and mechanism and carry profiles of causal, psychological, narrative and social continuity; prove that transmission and psychological continuity are logically independent, each holding in the other’s absence; and argue that the normative questions—consent, property, responsibility, deletion—are more informatively organised by transition type than by identity. The account is stated as three separable theses: that

closure is objective causal structure; that first-person anticipation tracks transmission rather than psychological similarity, a bridge principle defended abductively and flagged as the contested premise; and that once persistence architecture is a design variable it is an ethical and political one, deserving the scrutiny now given to a security model.

Keywords: personal identity; artificial consciousness; mind uploading; fission; causal processes; psychological continuity; mind duplication; process metaphysics; AI welfare; digital replicas.

Subject classification: Philosophy of Mind; Metaphysics; Ethics of Artificial Intelligence.

Contents

1	Introduction	4
1.1	A problem that was safe until it wasn't	4
1.2	The gap in the standard responses	4
1.3	The thesis	5
1.4	The relation to Parfit	6
1.5	What the paper does not claim	6
1.6	Plan	7
2	Conceptual preliminaries	7
2.1	Five questions that are not one question	7
2.2	Numerical and qualitative identity	8
2.3	States and histories	8
2.4	The subject-generating process	9
2.5	Terminology	9
3	Theories of persistence as a comparative test	10
3.1	Somatic and animalist accounts	10
3.2	Memory and psychological continuity	10
3.3	Closest-continuer and best-candidate theories	11
3.4	Four-dimensionalism and stage views	11
3.5	Narrative accounts	12
3.6	Parfitian reductionism	12
3.7	The nearest antecedents, and what is left to do	13
3.8	Where the theories agree, and why it matters	14
4	From causal processes to descriptive bottlenecks	14
4.1	The distinction from the philosophy of causation	14
4.2	What the backup case is, and is not	15
4.3	Closed records, dead intervals, and the transmission fraction	15
4.4	Mark transmission for subject-generating processes	17
4.5	The Continuity Criterion	17
4.6	Five properties of the criterion	18
4.7	Why this is not merely a verbal proposal	19
5	Applying the criterion	19
5.1	Cases where the criterion agrees with common sense	19
5.2	Migration cases	20
5.3	The virtual-machine migration cases	21
5.4	Backup and restoration	22
5.5	Fission	23
5.6	Summary of verdicts	23

6	Lineage graphs: what replaces identity	23
6.1	The structure	24
6.2	Formal properties	25
6.3	Profiles: persistence unbundled	26
6.4	Survival value as a graph measure	27
6.5	Fission without paradox	27
6.6	The practical payoff: typed, profiled edges organise the normative questions	28
7	The first-person problem	29
7.1	The question that will not reduce	29
7.2	Three theses	29
7.3	What the descriptive thesis supplies	30
7.4	The abductive case for the bridge	30
7.5	The condemned-duplicate test	31
7.6	Anticipation as the relevant attitude	32
8	The locus of identity in artificial systems	32
8.1	Candidate loci	33
8.2	The stateless-inference result	33
8.3	The KV-cache distinction	35
8.4	Forking is the default, not the exception	35
9	Objections and replies	35
10	Persistence architecture as a design variable	39
10.1	From metaphysics to specification	39
10.2	What a persistence specification would state	40
10.3	Consent problems with no biological analogue	40
10.4	The political dimension	41
10.5	A right to continue, and its limits	42
11	A research programme, and what would refute the account	42
11.1	Open problems	42
11.2	What would refute the account	43
11.3	Conclusion	43
	References	44

1 Introduction

1.1 A problem that was safe until it wasn't

Philosophical theories of personal identity were developed under a set of background conditions so stable that they were rarely stated. A person occupies one body. That body houses one brain. The brain has one causal history, running backwards without branches. It cannot be duplicated, cannot be restored from an earlier state, cannot be paused and resumed on different hardware, and cannot be merged with another person after a period of separate life. Under those conditions, a great many candidate criteria for personal persistence—bodily continuity, brain continuity, psychological connectedness, autobiographical memory, narrative self-interpretation, social recognition, legal personhood—deliver the same verdict about the same cases, because in human life they travel together.

The philosophical tradition has long known that they are separable in principle. Locke (1694) separated memory from bodily continuity; Butler (1736) and Reid (1785) pressed the circularity and transitivity objections; Williams (1970) showed how differently the same procedure reads depending on whether it is described in psychological or bodily terms; Parfit (1971, 1984) made fission the organising case for the whole field. But the separation was always achieved by stipulation. The thought experiments were, as Wilkes (1988) complained, doing metaphysics with machinery that no one could build.

That complaint is losing its force. Not because anyone has built a conscious machine—no one has, and nothing here assumes otherwise—but because the operations the thought experiments stipulate are ordinary in computing. Checkpoint and restore, process forking, live migration between hosts, snapshotting, rollback and state merging are not exotic; they are the standard vocabulary of systems engineering (Clark et al., 2005). If a conscious system is ever realised on such infrastructure, the separations that Locke and Parfit had to imagine will be available as API calls. The question then stops being what *would* follow if minds were copyable and becomes what *does* follow, for a system whose persistence semantics are decided in a design review.

This paper is about what should be decided in that design review, and about the metaphysics needed to decide it.

1.2 The gap in the standard responses

Two positions dominate contemporary discussion of copying.

The first is psychological continuity theory: a later person is the earlier one if the later mental states are appropriately caused by the earlier ones, usually with a non-branching clause appended to preserve the logic of identity (Shoemaker, 1984; Parfit, 1984; Noonan, 2003). The second is Parfitian reductionism: fission shows that psychological continuity can hold one-to-many while identity cannot, and rather than patch identity we should conclude that identity is not what matters in survival (Parfit, 1971, 1984).

The two positions disagree about a great deal. But they agree about something that this paper denies. Both treat the causal route by which a later mental state was produced as mattering only instrumentally—as a way of ensuring that the right psychological content arrives. Parfit is explicit that the cause may be “any cause” (Parfit, 1984, p. 207). On both views, once you fix that the later system has the memories, the intentions, the character and the projects, and that these were produced in a suitably dependable way by the earlier system, there is nothing further of prudential significance to settle. The teletransporter, on this shared assumption, is at worst a strange way to travel.

I think this is a mistake, and that the mistake is diagnosable. Psychological continuity is a relation between *informational contents*. It is preserved whenever the content arrives, no matter how. That makes it indifferent to a distinction which physical science has independent reason to draw: the distinction between a process that carries its own history forward and a sequence of states that merely looks continuous because something outside it is generating each stage in turn.

Salmon (1984) drew that distinction for physics. A causal process—a moving ball, a light pulse—can transmit a mark: scratch it here and it stays scratched there. A pseudo-process—the shadow of the ball, the bright spot a rotating beacon sweeps across a dome—cannot. The spot on the dome moves continuously and may move faster than light, but a local modification made to it at one place does not persist to the next, because the spot at each instant is generated afresh by the beacon rather than produced by the spot that preceded it. Continuity of appearance is not continuity of process.

The central claim of this paper is that this distinction is exactly the one missing from the personal identity debate, and that when it is supplied, the copying cases sort themselves out.

1.3 The thesis

A mind restored from a stored record is causally descended from its original without being dynamically generated by it. After the record is written, the original's ongoing life makes no difference to the successor: the continuity between them is real as continuity of content and absent as continuity of generation.

I make this precise in Section 4. The load-bearing notion is the *closed record*: a physical state, produced from *A*, whose content has ceased to depend on *A*'s further evolution, and from which a successor *B*'s constitutive state is initialized across a *dead interval*—a period in which that state is neither open to intervention on *A* nor in use by *B*. The notion is defined causally throughout; nothing turns on the record *representing* anything. The resulting criterion is graded: a transition transmits some fraction τ of the successor's constitutive state and re-instantiates the remainder through closed records, with continuation ($\tau = 1$) and reconstruction ($\tau = 0$) as the endpoints.

The account then divides into three theses of different strength (Section 7.2): that this structure is objective causal fact, defended outright; that first-person anticipation tracks it, a bridge principle defended abductively and flagged as the paper's contested premise; and that the practical questions of copying are better organised by transition type than by identity, defended by exhibition.

Three features of this criterion are worth flagging at the outset, because they distinguish it from positions it might be confused with.

It is not physical continuity in disguise. The criterion permits complete material replacement. Gradual prosthetic substitution of neural tissue, carried out while the system runs, transmits marks and involves no descriptive bottleneck, and therefore preserves a subject on this account even though no original matter survives. Conversely, the criterion *rejects* some procedures that are physically continuous in the ordinary sense: pausing a brain, scanning it region by region, and rebuilding it in place from the scan fails, notwithstanding that it is gradual and that the rebuilt system occupies the same location. What matters is not the persistence of

matter but whether the system’s own dynamics, rather than a representation of them, generate the succession.

It is not a soul. Nothing in the criterion appeals to anything beyond physical states and their causal relations. Screening off is a notion from causal modelling (Reichenbach, 1956; Pearl, 2009); mark transmission is a notion from the physics of causation (Salmon, 1984; Dowe, 2000). The fact the criterion tracks is a historical and relational fact, not an intrinsic one—but so is being a widow, and no one thinks widowhood requires a nonphysical substance.

It does not save numerical identity. Applied to Parfit’s fission case, the criterion says that *both* resulting hemispheres continue the original, since each transmits marks and neither passes through a bottleneck. Continuation is therefore one-to-many and so is not identity. I take this to be the right result rather than a defeat, and Section 6 builds the consequence into a formalism rather than papering over it.

1.4 The relation to Parfit

It follows that the position defended here is a revision of Parfit rather than a refutation of him. Parfit’s central negative claim—that numerical identity is not what matters in survival—I accept, and Section 6 gives it a formal statement. What I reject is his positive identification of what does matter with psychological continuity and connectedness.

The disagreement can be put sharply. Psychological continuity survives descriptive bottlenecks: a system rebuilt from a checkpoint has, by construction, all the memories and intentions of the system checkpointed. If psychological continuity is what matters, then a person facing destruction and subsequent reconstruction has nothing prudentially significant to fear, and the enormous intuitive resistance most people feel to that conclusion is simply an error. Parfit accepted this consequence. I argue in Section 7 that the resistance is instead tracking something real, and that what it tracks is the bottleneck. The intuition is not a superstition about souls; it is a rough detection of the difference between transmission and re-instantiation.

That gives the account a testable shape, in the modest sense available to philosophy. Bourget and Chalmers (2023) report that the profession is close to evenly divided on whether teletransportation is survival or death. A theory that explains *why* the case divides people, and predicts which nearby cases should divide them similarly and which should not, is doing more than a theory that assigns a verdict and calls the other side confused. Section 5 sets out a series of cases, including several the literature has not discussed, on which the criterion makes predictions that psychological continuity theory does not.

1.5 What the paper does not claim

Three disclaimers, since the topic invites overreach.

First, nothing here shows that artificial systems are or will be conscious. The criterion is conditional throughout: *if* a system is subject-generating, here is when a later system continues it. Whether current or future systems satisfy the antecedent is an empirical and theoretical question this paper does not settle, and on which I take the responsible position to be the precautionary one recommended by Butlin et al. (2023) and Long et al. (2024).

Second, the criterion does not entail that copies are worthless, that reconstruction is impermissible, or that a replica has no claims. A reconstruction may be a conscious being with full

moral status, an heir to a person's projects, and a legitimate participant in their relationships, while not being a continuation of them. Section 10 argues that most of the practical questions are better indexed to the type of transition than to the identity verdict, which is one reason the identity verdict has been made to carry more weight than it can bear.

Third, I do not claim the criterion is complete. It says when a transition breaks a subject. It does not say what makes a system subject-generating in the first place, and it is deliberately neutral among theories of consciousness (Baars, 1988; Tononi, 2008; Dehaene et al., 2017; Seth, 2025). A full account would need both halves. This paper supplies the half that the copying cases actually turn on.

1.6 Plan

Section 2 separates five questions that discussions of copying routinely run together, and states the state/history distinction on which everything later depends. Section 3 reviews the main theories of persistence as a comparative test rather than a menu, locates the assumption the paper attacks in the tradition that actually holds it, and situates the account against its nearest antecedents. Section 4 is the core: it takes the process/pseudo-process distinction as motivation, shows why it does not settle the backup case, and builds the notions—closure, dead intervals, the transmission fraction—that do, culminating in a graded Continuity Criterion. Section 5 applies it to fourteen cases, including live virtual-machine migration, and extracts the criterion's non-obvious predictions. Section 6 develops the lineage-graph formalism, proves that transmission and psychological continuity are logically independent, and generalises both into a profile of continuity relations. Section 7 separates the account into descriptive, bridge and normative theses, and is candid that the bridge is defended only abductively. Section 8 applies the criterion to the architecture of systems that exist now, and derives the locus-conditional result that under transcript-replay execution a conversation is a succession of reconstructions. Section 9 answers ten objections and is explicit about which replies are conclusive and which merely price the commitments. Section 10 turns to persistence architecture as an ethical and political design variable. Section 11 sets out a research programme.

2 Conceptual preliminaries

2.1 Five questions that are not one question

Discussions of copying go wrong early and often because a single case is made to answer several questions at once. Following and extending the taxonomy in Olson (2007) and the framing of Olson (2023), I separate five.

- (Q1) **The consciousness question.** Is the successor a subject of experience at all?
- (Q2) **The psychological question.** Does the successor have the source's memories, character, intentions, values and projects?
- (Q3) **The identity question.** Is the successor numerically the same subject as the source?
- (Q4) **The prudential question.** What, if anything, should the source care about in the successor's future?
- (Q5) **The normative question.** What is owed to the successor, and by whom, and what may be done to it?

These are logically independent in both directions. (Q1) can be answered yes and (Q3) no: a copy may be fully conscious and yet a different subject. (Q2) can be answered yes and

(Q1) no: a system might reproduce all the functional roles of memory and character without phenomenal experience, which is precisely what is at issue between functionalists and their opponents (Chalmers, 1996; Seth, 2025). (Q3) can be answered no and (Q5) yes, emphatically: a reconstruction that is not the original may nonetheless be a person with the full complement of rights, as Schwitzgebel and Garza (2015) argue for artificial systems generally. And (Q4) is not settled by (Q3), which is Parfit’s principal insight.

Most of the apparent deadlock in the copying literature comes from treating a verdict on one of these as a verdict on the others. A common pattern: opponents of uploading argue that the upload is not *you* (a claim about Q3) and are answered with the observation that the upload would be conscious and would have your memories (claims about Q1 and Q2), which is not a response. Another: defenders of digital replicas argue that a replica has moral standing (Q5) and are taken to have argued that it is the deceased person (Q3), which it is not. Karpus and Strasser (2025) are unusually careful about this, and their proposal that a replica may *extend* a person’s personhood without being identical with them is best read as an answer to Q5 that is deliberately silent on Q3.

This paper is principally about (Q3) and (Q4), and argues that they come apart, but in a different place from where Parfit put the break.

2.2 Numerical and qualitative identity

The distinction is old and still load-bearing. Two objects are qualitatively identical when they share their properties; a thing is numerically identical with itself and with nothing else. Numerical identity is an equivalence relation: reflexive, symmetric, transitive. It is also, crucially, *one-to-one*: if $a = b$ and $a = c$ then $b = c$.

That last feature is what fission breaks, and it is worth being precise about why. Suppose a procedure yields B and C from A , and suppose the relation R that we had been treating as constitutive of identity holds between A and B and between A and C . Then either R is not identity, or $B = C$, which is false since B and C are in different rooms having different thoughts. No amount of insistence that B “really is” A survives the observation that C has an equally good claim, established by the same relation, and that the two claims cannot both be honoured.

Three responses are available and all have been taken. Add a non-branching clause to R , so that R holds only when there is no rival (Shoemaker, 1984). Select a winner by a comparative measure (Nozick, 1981). Or conclude that R is not identity and that identity was the wrong thing to be tracking (Parfit, 1984). Section 3 takes up the costs of each; Section 6 takes the third road while disagreeing with Parfit about what replaces identity.

2.3 States and histories

The distinction on which the whole paper turns can be stated in two lines.

Let $S_t(P)$ denote the complete physical state of a system P at time t , specified to whatever grain the correct theory of consciousness cares about. Let $H_t(P)$ denote the causal history by which that state came to obtain.

A copy is a system B such that

$$S_t(B) = S_t(A) \quad \text{while} \quad H_t(B) \neq H_t(A).$$

The question is whether personal persistence supervenes on the state or on the history. If on the state, the copy is the original, and the metaphysics of copying is trivial. If partly on

the history, the copy may be a successor without being the original, and everything depends on which histories count.

It is worth resisting a tempting dismissal here. One might say: two systems with identical intrinsic properties cannot differ in any respect that matters, so history is idle. But this is false quite generally and not only for persons. Two hurricanes with the same instantaneous wind field are two hurricanes, not one, and which one is which is settled by their histories. Two banknotes with the same serial number differ in that one is counterfeit; the difference is entirely historical and entirely real. Two chemically indistinguishable diamonds differ in that one was mined and one was synthesised. In each case the intrinsic description is exhausted and a genuine distinction remains.

The objection worth taking seriously is not that historical facts are unreal but that they are the *wrong sort* of fact to determine the fate of a subject. I take that objection up in Objection 9, where I regard it as the strongest thing that can be said against the position.

2.4 The subject-generating process

I will speak throughout of a *subject-generating process*. The phrase is meant to be a placeholder and to be recognised as one.

Definition 2.1 (Subject-generating process). A physical process P is *subject-generating* over an interval $[t_1, t_2]$ iff, throughout that interval, P realises whatever dynamics are in fact sufficient for phenomenal consciousness.

This is deliberately theory-neutral. If the global workspace account is right (Baars, 1988; Dehaene et al., 2017), the subject-generating process is whatever implements broadcast and ignition. If integrated information theory is right (Tononi, 2008), it is whatever realises the relevant cause–effect structure. If Seth (2025) is right that consciousness is tied to the self-maintaining dynamics of living systems, the class may be narrower than functionalists suppose, and may exclude digital systems entirely.

The criterion developed here is compatible with all of these, and that is the point. It is a criterion about *persistence*, not about *constitution*. It presupposes that some class of processes generates subjects and asks when a later member of that class continues an earlier one. An account of persistence that had to wait for a settled theory of consciousness would never be given, and the engineering decisions it bears on are being made now.

Two consequences follow immediately. First, if no digital system is subject-generating, the criterion is vacuously satisfied for all digital cases and the whole discussion is about biological persons and hypothetical machines. That is a coherent position and the criterion does not beg the question against it. Second, if some digital systems are subject-generating, the criterion applies to them without modification, since nothing in it mentions substrate.

2.5 Terminology

I reserve the following words and will use them strictly.

- **Continuation.** The relation defined by Criterion 4.4 at $\tau = 1$. Written $A \triangleright B$ (“ B continues A ”).
- **Reconstruction.** A successor produced through a total descriptive bottleneck ($\tau = 0$). Written $A \dashrightarrow B$.
- **Successor.** Anything standing in either relation; the neutral term.

- **Branch.** One of several continuations of a single source.
- **Replica.** A system built to resemble a source, with no commitment about either relation. Danaher and Nyholm (2024a,b) use “duplicate” for roughly this.
- **Lineage.** The full ancestry of a process under continuation and reconstruction together, with edge types recorded (Section 6).

The word I will avoid, except when reporting others’ views, is *survival*. It is systematically ambiguous between the identity question and the prudential question, and the ambiguity is exactly where the argument lives.

3 Theories of persistence as a comparative test

The theories reviewed here are not a menu of alternatives to be sampled. They are a set of instruments, each of which measures something real, and copying is the case that makes them disagree. My interest is in locating precisely where the disagreement lies, because that is where the argument of Section 4 enters.

3.1 Somatic and animalist accounts

Bodily and animalist theories hold that we are biological organisms and persist under organismic persistence conditions (Olson, 1997, 2007). The view is unusually clear about copying: a scan-and-build procedure does not move an organism, so the product is not the person, however psychologically rich it is. Animalism thereby avoids the fission problem for copying entirely, since copying a psychology does not copy an animal.

Its costs are equally clear. It makes brain transplantation into the death of the donor and the acquisition of memories by the recipient, which is difficult to accept. Johnston (1987) argues that the choice between animalist and psychological criteria is less forced than it appears, since our concept of a human being may already be a hybrid; Wiggins (1967) had earlier made spatio-temporal continuity under a sortal the organising notion, which generalises awkwardly to substrates that admit duplication. It has nothing to say about a conscious artificial subject: if persistence conditions are organismic, and a machine is not an organism, then either machines cannot be subjects or their persistence conditions must be supplied from elsewhere. Olson (2007) is candid that personal ontology is prior to personal persistence, and animalism’s answer to the ontological question rules out the cases this paper is about.

What animalism gets right, and what I want to preserve, is its insistence that psychological facts do not float free of the physical thing that has them. The criterion of Section 4 keeps that insistence while dropping the biological restriction.

3.2 Memory and psychological continuity

Locke (1694) tied personal identity to continuity of consciousness, and the objections arrived promptly: Butler (1736) that memory presupposes identity rather than constituting it, and Reid (1785) that the relation is not transitive, since the general remembers the officer who remembers the flogged boy while the general does not remember the boy. Hume (1739) pressed a deeper worry that bears directly on the present argument: introspection discloses perceptions but no perceiving self among them, so that whatever unity a mind has must be constituted by relations among its states rather than found as a further item. A process account is one way of taking that conclusion seriously without concluding that the unity is fictitious.

The modern repairs are well known. Replace direct memory with overlapping chains of psychological connection (Shoemaker, 1984; Parfit, 1984); replace “remembers” with a causally constrained notion of quasi-memory to dodge the circularity; add a non-branching clause to preserve one-to-one-ness. The resulting theory handles gradual change gracefully and matches ordinary judgement in nearly all actual cases.

Two features of it matter here. The first is the appropriate-cause condition. Psychological continuity requires not merely that the later states resemble the earlier ones but that they were caused by them. This looks like it does the work I want. It does not, and the reason is Parfit’s own liberality about the cause: he explicitly allows that the cause may be of any kind, on the grounds that insisting on the normal cause is arbitrary (Parfit, 1984, pp. 204–209). Once “any cause” is permitted, a reconstruction from a scan qualifies. The appropriate-cause condition, as standardly developed, filters out coincidental resemblance but not re-instantiation from a description. That is the gap.

The second is the non-branching clause, which is widely and rightly felt to be *ad hoc*. If *B* would have been the original had *C* not been built, it is very hard to see how building *C*—perhaps on another continent, perhaps without *B*’s knowledge—can retroactively unmake *B*’s claim. This is the branching-interference objection, and no version of the clause has answered it satisfactorily.

3.3 Closest-continuer and best-candidate theories

Nozick (1981) proposes that a person is continued by whichever candidate is closest, provided it is close enough and no rival ties. This preserves identity in the ordinary case and delivers a verdict in most branching ones.

The difficulties are structural rather than incidental. The theory makes identity extrinsic: whether *B* is *A* depends on facts about *C*, which is the branching-interference objection again in a more explicit form. It requires a closeness metric over which nothing but intuition adjudicates. And in the symmetric case, where the candidates tie, it delivers no verdict at all, which is the case that motivated the whole discussion.

3.4 Four-dimensionalism and stage views

If persons are spacetime worms composed of temporal parts (Sider, 2001), or momentary stages standing in counterpart relations (Sider, 1996), fission can be redescribed rather than resolved. Lewis (1976) argues that two persons were present all along, sharing their pre-fission stages, so that psychological continuity and identity can both be preserved; the cost is that before the division there were two people in the room, and that ordinary counting was wrong.

Sider (2018) adds a consideration that bears directly on copying: the relations underwriting identification can be asymmetric. His case involves a daytime subject and a nighttime subject where the later remembers the earlier but the earlier neither remembers nor anticipates the later. The application to copying is immediate and, I think, underappreciated. After a scan, the reconstruction looks back and finds a full autobiography; the original looks forward and finds—what? The retrospective relation and the prospective relation are not the same relation, and the copying literature has largely argued about the first while caring about the second.

Four-dimensionalism is a real option and I do not argue against it here. I note only that it is a thesis about the ontology of persistence and does not by itself say which cross-temporal

relation constitutes the unity of a worm. That question is the one this paper answers, and a four-dimensionalist can accept the answer.

3.5 Narrative accounts

Schechtman (1996, 2014) hold that persons constitute themselves through the narratives they construct and inhabit. The view illuminates a great deal about what it is to have a life rather than merely a history, and Schechtman (2014) in particular is careful to distinguish the questions that narrative answers from those it does not.

For copying, narrative accounts have a characteristic strength and a characteristic limit. The strength: a reconstruction genuinely inherits a narrative, can continue its themes, and can be received by others as continuing it, which is a real form of persistence and explains much of what people want from digital legacy systems. The limit: two branches can inherit one pre-fission story and then author two incompatible continuations of it, so narrative inheritance is one-to-many exactly where identity needs one-to-one. Narrative grounds social and practical continuity without grounding numerical identity, and I think its proponents can accept this.

3.6 Parfitian reductionism

Parfit (1971, 1984) argues that once we have described the physical and psychological facts and the relations among them, there is no further fact about identity to discover; that identity is therefore not what matters; and that what matters is psychological continuity and connectedness, which can hold one-to-many. Fission is not a puzzle to be solved but a demonstration that we were asking the wrong question.

The position is powerful and I accept its negative half. Two recent lines of resistance are worth recording.

Ehring (2018) challenges the inference from fission to ordinary survival. Parfit reasons that where identity and psychological continuity diverge, continuity is what matters; therefore identity never matters. Ehring observes that this does not follow: what matters in the divergent case may be overdetermined in the convergent one, so that identity makes an independent contribution in ordinary survival even though continuity carries the weight under fission. The point is well taken and I make use of it, though my own route is different: I do not argue that identity matters independently, but that the relation Parfit identified with what matters is not fine-grained enough.

Goldwater (2021) presses a transmission problem: uploading may reproduce the information constituting a person without transmitting the person. His route to that conclusion is quite different from mine and should not be conflated with it. Goldwater argues that uploading is possible only if a mind is an abstract object rather than a concrete particular—uploadable in the way a book is uploadable and a painting is not—and that since the leading theories of personal identity treat minds as concrete particulars, uploading is incompatible with them. That is an argument from the type/token distinction. Mine is an argument from causal process structure, and it does not require minds to be concrete particulars in Goldwater's sense; it requires only that subject-generating processes be individuated by their histories, which is compatible with a range of views about their ontological category. Weber (2025) presses a related multiplicity objection, that the possibility of many simultaneous uploads undermines the inference from psychological similarity to identity. Cerullo (2015) and Corabi and Schneider (2012) raise structurally similar

concerns for branching and for the metaphysics of uploading respectively, and Walker (2011) surveys the options for an uploading-friendly criterion.

Dainton and Bayne (2005) argue that phenomenal continuity, rather than psychological continuity in the Lockean sense, is the guide to persistence. I agree that something other than psychological continuity is wanted, but locate it in the generating process rather than in the stream of experience, for the reason given in Case 2: experience lapses nightly and the process does not.

3.7 The nearest antecedents, and what is left to do

Three positions come close enough to the present account that the difference should be stated before the argument rather than after it.

Russell's causal-line theory. On a broadly Russellian view, a person's persistence consists in successive stages belonging to a single causal line. This is the correct genus, and the present account is a species of it. But a causal-line theory as usually stated does not distinguish among causal lines, and the copying cases require exactly that distinction: a scan, a transmission and a reconstruction form a perfectly good causal line, and the question is whether it is the right kind. Section 4 is an attempt to say which causal lines carry a subject, and this is what a bare causal-line theory leaves open.

Unger's physically based approach. Unger (1990) is the closest antecedent I know of. He argues that contemporary philosophers have underrated the importance of physical continuity to survival, and defends an account that is psychologically aimed but physically based: what matters is the persistence of the physical basis realising the capacity for consciousness, not the psychology it supports. I agree with the direction and disagree about the criterion. Unger's requirement is naturally read as demanding continuity of the realising *stuff*, which makes the account hostage to questions about how much matter may be replaced how quickly. The present criterion places no requirement on matter at all: it permits complete material replacement (Case 5) and refuses some materially conservative procedures (Case 6). What it requires is transmission, which is a property of the process rather than of its constituents. That difference is what makes the account portable to substrates Unger was not writing about.

Anti-pattern arguments. A line of recent work argues from duplication directly against pattern and functional theories, concluding that identity may be primitive or not further analysable. I share the negative conclusion that pattern preservation is insufficient and reject the positive one. The whole point of Section 4 is that the residue left over when pattern is subtracted is not primitive: it is mark transmission, which is a physical relation with an established analysis in another subfield. An account that ends in primitivism has given up exactly where the work begins.

The contribution claimed here is therefore narrow and should be stated narrowly. It is not a new position about what matters in survival; the position is recognisably Unger's, and the sceptical intuition it systematises is very old. It is a criterion—a statement of which causal relations constitute continuation, drawn from an existing account of physical causation, sharp enough to sort cases that the received views cannot sort.

3.8 Where the theories agree, and why it matters

Set animalism aside, since it declines the question rather than answering it. Set aside also the history-constitutive minority just surveyed in Section 3.7—the causal-line view, Unger’s physically based account, and those four-dimensionalisms on which the structure of the whole worm is constitutive. What remains is the tradition that has actually governed the copying debate: psychological continuity theory and its Parfitian development. That tradition shares a commitment that can be stated precisely.

Principle 3.1 (State sufficiency). If the later system has the right psychological content, and that content was produced by a dependable causal route from the earlier system, then whatever personal relation is constitutive obtains. Among dependable routes, the particular manner of the route is not itself constitutive.

Parfit’s “any cause” liberality (Parfit, 1984, pp. 204–209) is the explicit form of this commitment, and the non-branching clause is its characteristic patch. It is what makes the tradition return the same verdict on the teletransporter that it returns on ordinary survival, and what makes the copying cases feel, to many readers, as though something is being missed that they cannot name.

Principle 3.1 is the target of this paper—not because every theory of persistence accepts it, but because the theories that dominate the copying literature do. I argue that the manner of the causal route is constitutive, that a principled and non-arbitrary line can be drawn within the class of dependable routes, and that the materials for drawing it lie in the philosophy of physical causation—though the drawing itself requires an extension that Section 4 supplies.

4 From causal processes to descriptive bottlenecks

4.1 The distinction from the philosophy of causation

Salmon (1984), developing a suggestion of Reichenbach (1956), distinguishes causal processes from pseudo-processes. A causal process is a world-line capable of *transmitting a mark*: introduce a local modification at one point and, absent further intervention, the modification persists to later points. A moving ball scratched at t_1 is scratched at t_2 . A pseudo-process is a spatiotemporally continuous sequence that cannot do this.

The canonical example is a rotating beacon inside a circular wall. The bright spot travels around the wall continuously, and if the wall is large enough and the beacon fast enough, the spot moves faster than light without violating relativity. Now place a red filter at one point on the wall. The spot turns red as it passes and immediately reverts. The modification does not travel with the spot, because the spot at each moment is not produced by the spot at the previous moment. It is produced by the beacon. The apparent motion is an artefact of a succession of distinct events with a common cause.

Dowe (2000) refines Salmon’s account, replacing mark transmission with the possession of a conserved quantity, partly to avoid difficulties with counterfactual formulations. The refinements matter for the metaphysics of causation and matter less here; what I need is the structural contrast, which both accounts preserve. Continuity of appearance is not continuity of process. A sequence can be smooth, unbroken, and fully law-governed, and still fail to be a thing that carries itself forward.

The proposal of this paper is that this contrast, properly extended, is what the personal identity debate has been missing. The extension is not automatic, and seeing exactly where it fails to be automatic is where the paper’s own contribution begins.

4.2 What the backup case is, and is not

Consider a system A checkpointed at t_1 to an archive, run on until t_2 , destroyed, and rebuilt at t_3 from the archive as B.

It is tempting to summarise what follows as: B stands to A exactly as the later spot-positions stand to the earlier ones—continuity of appearance without continuity of process. The summary must be resisted, because it is false, and the way it is false is instructive.

The chain from A through the archive to B is an entirely ordinary causal chain. A produced the archive; the archive produced B; a mark introduced into A *before* t_1 propagates through the archive into B, and a modification of the archive itself alters B. Nothing here resembles the spotlight, whose successive spot-stages do not produce one another at all but are each generated independently by the beacon. In the backup case there is genuine production at every link. An analysis that denies this denies an evident causal fact—and Salmon’s dichotomy, applied mechanically, classifies the backup chain as a causal process and is done.

What the dichotomy misses is a distinction *within* causal descent. Everything A did after t_1 —every thought, decision, revision, injury—makes no difference whatever to B. B depends completely on A-as-of- t_1 and not at all on A’s subsequent life. A went on generating states for another interval, and none of that generation reaches B; what reaches B is what the archive holds, and the archive stopped listening at t_1 .

The backup case is therefore a third structure, which Salmon’s dichotomy does not contain: *causal descent without dynamical generation*. B is descended from A; B is not generated by A’s ongoing activity. What Salmon supplies is the lesson that continuity of appearance and continuity of generation are different things, together with a method—counterfactual probing of dependence—for telling them apart. The concept that captures the third structure has to be built, and building it is the business of this section. It owes Salmon its motivation, not its content.

4.3 Closed records, dead intervals, and the transmission fraction

One preliminary, because the informal word “description” is dangerous. An archive does not intrinsically announce that it is a description. It is a physical state that some mechanism can use to produce another physical state; whether anyone would call it a representation is neither here nor there. If the account rested on a semantic notion—a state that *represents* the source—it would inherit observer-relativity, since representation is cheap and ubiquitous. Everything that follows is therefore defined in causal vocabulary, using the interventionist apparatus standard in causal modelling (Reichenbach, 1956; Pearl, 2009). “Descriptive bottleneck” survives as a label; the load-bearing notions are *openness*, *closure*, and the *dead interval*.

Definition 4.1 (Openness and closure). Let x be a component of a prospective successor’s initial constitutive state. x is *open* to the source at time t iff an intervention on the source token at t would make a difference to x ’s value or to x ’s subsequent effect on the successor’s evolution. x *closes* at the first time after which it is never again open.

Definition 4.2 (Dead intervals and the transmission fraction). Let the source’s *generative endpoint* be the last time its subject-generating dynamics run, and let x ’s *uptake* be the first time x is in use by the successor’s subject-generating dynamics.

- (i) A *closed record* for x is a physical state, causally produced from the source, which fixes x ’s initial value and which has closed in the sense of Definition 4.1 while remaining causally upstream of x .

- (ii) x crosses a *dead interval* iff there is a period after x closes and before x 's uptake during which x is neither open to the source, nor in generative use, nor in *causal transit*—where x is in causal transit while its value is borne by an ongoing propagation that terminates directly in generative uptake, with no intervening state whose persistence is independent of progress toward that uptake, and *stored* while it rests in a state whose persistence is independent of any progress toward uptake.
- (iii) x is *transmitted* iff it reaches uptake without crossing a dead interval; x is *re-instantiated* iff it crosses one.
- (iv) The *transmission fraction* $\tau \in [0, 1]$ of a transition is the proportion of the successor's initial constitutive state, under the measure appropriate to the subject-generating process, that is transmitted. The transition passes through a *total descriptive bottleneck* iff $\tau = 0$, a *partial* one iff $0 < \tau < 1$, and none iff $\tau = 1$.

Five remarks. First, staticity—the intuitive requirement that a record be *finished*—is now a consequence rather than a stipulation: a record that continues to track the source has not closed, so a continuously updated mirror is no bottleneck. Second, the definitions are token-level and counterfactual throughout: whether x closed, and when, is settled by what interventions on the source *would have* done, not by what happened—a component no actual write touched may have remained open the whole time, because a write would have propagated. Third, severance is structural, not temporal. A dead interval of a nanosecond and a dead interval of a century classify alike; what matters is that the generative chain was severed, not for how long. I return to whether this is a cost in Section 11. Fourth, the measure over constitutive components is not a byte count. What τ weights is fixed by the theory of the subject-generating process, and the relevant weighting is by causal role in the generative dynamics, not by amount of state: a component can be minute and indispensable, and a large store can be idle. Where this paper writes $\tau \approx 0$ or $\tau \approx 1$, the claim is that the classification is insensitive to any reasonable choice of measure; where it is sensitive—the execution-context residue of Case 10 is the live example—the paper says so and leaves the weighting to the theory that owes it (Section 11).

Fifth, the transit clause in (ii) is load-bearing, and it answers what would otherwise be the sharpest objection to the whole apparatus. Ordinary physical signalling contains gaps. A neurotransmitter that has left the terminal but not yet bound, a photon between components, a pulse mid-wire, a packet in flight: in each case influence has left the source and not reached the recipient, so for an interval the carrier is neither open nor in use, and a definition without the transit clause would classify every synapse as a bottleneck and ordinary transmission as serial re-instantiation. The clause draws the needed line causally: influence *travelling*—borne by a propagation that terminates directly in generative uptake—is not a dead interval; a record *waiting*—resting in a state that persists whether or not uptake ever comes, read, if at all, by an exogenous process—is. The neurotransmitter's binding is the terminus of the very diffusion that bears it; the checkpoint's persistence owes nothing to any progress toward its reading.

Note what the clause does *not* exempt, and why. The teleporter's beam is propagation, but its contents come to rest in a not-yet-running assembly—a state whose persistence owes nothing to progress toward uptake—and rest is storage. What stops the beam is the rest, not any requirement that the receiver antecedently be running: a propagation may itself ignite the dynamics it feeds, as when a volley recruits a quiescent circuit into activity, and such ignition is transit in good standing. The exemption generalises: the gaps the clause exempts are those across which generative activity is handed on without ever merely waiting, which is the paper's thesis arrived at from below. It also yields a prediction worth owning rather than avoiding: a streaming proce-

ture whose arriving signal directly and continuously drives a successor's dynamics into operation, never forming a resting record, would preserve substantial causal continuation—teleportation become gradual replacement at a distance, the verdict tracking the mechanism and not the name. Two grain-level safeguards stack on top. Micro-carriers are not constitutive components (fourth remark), so synaptic traffic never takes a constitutive component out of use in a running brain; and where a whole constitutive component is itself in flight—a faulted page in post-copy migration—it passes directly into generative use, and the clause classifies it as transit. Mixed and marginal cases remain. A procedure in which some constitutive components stay in continuous transit while others enter storage yields an intermediate τ ; and a single component's episode can itself be borderline—a packet held at a store-and-forward relay is neither clearly propagating nor clearly at rest—in which case it is the classification of that component, not the definition of τ , that goes indeterminate. I flag the borderline rather than legislate it.

The backup case now has its analysis. Every component of B's initial state closed at t_1 and sat in a closed record across a dead interval until t_3 : $\tau = 0$, a total bottleneck, re-instantiation of everything. Ordinary life is the opposite pole: nothing closes, $\tau = 1$. Between the poles lies a continuum, and the cases of Section 5 populate it.

4.4 Mark transmission for subject-generating processes

Definition 4.3 (Mark transmission). A process P *transmits marks* over $[t_1, t_2]$ iff for a suitably broad class of local modifications m applied at any $t \in [t_1, t_2]$, the effect of m persists in P's state at later times in $[t_1, t_2]$ without further intervention.

Two clarifications forestall obvious misreadings. First, “suitably broad” does not mean universal: a modification that destroys the system does not persist, and systems have error-correcting dynamics that wash out small perturbations. The condition is that the process be the sort of thing modifications can stick to, not that every modification stick. Second, mark transmission and openness are complementary probes of one structure: transmission asks whether influence propagates forward along the carried chain; openness asks whether the successor's components remain sensitive to the source until use. A transition transmitted in the sense of Definition 4.2 supports both, and a dead interval is precisely a stretch where both fail.

4.5 The Continuity Criterion

Criterion 4.4 (Continuity, graded form). Let A and B be processes. Then B *continues A to degree* τ iff:

(C1) Subject condition

both A and B are subject-generating (Definition 2.1);

(C2) Transmission condition

the transmitted portion of B's initial constitutive state reaches it through chains that transmit marks (Definition 4.3) without lapse;

(C3) Closure condition

τ is the transition's transmission fraction (Definition 4.2).

B is a *continuation* of A, written $A \triangleright B$, when $\tau = 1$; a *reconstruction*, written $A \dashrightarrow B$, when $\tau = 0$; a *partial continuation* for intermediate τ . The binary vocabulary used elsewhere in

this paper names the endpoints, and every binary verdict in Section 5 should be read as an endpoint classification under this graded form. For transitions near but not at an endpoint I write *continuation-dominant* ($\tau \approx 1$) and *reconstruction-dominant* ($\tau \approx 0$): these name partial continuations, not endpoint membership, and the approximation claims dominance under any reasonable constitutive measure (fourth remark, Section 4.3).

The three conditions do distinct work and none is redundant.

(C1) prevents absurd permissiveness. My writing this paper transmits marks to the paper and nothing closes en route, but the paper is not a subject and so is not my continuation. The criterion says when a *subject* continues, and presupposes an independent account of what makes a process subject-generating. (C1) also fixes the measure over which τ is assessed: what counts as the constitutive state, and how its components are weighted, is inherited from the theory of the subject-generating process—and Section 8 shows that this inheritance is where the hardest questions now live.

(C2) rules out transitions in which each stage depends on the last but nothing is carried—a system repeatedly rebuilt to a specification by an external controller, even a specification updated continuously.

(C3) is the novel condition and does the work in the copying cases. It is what separates transmission from re-instantiation, and it is what Principle 3.1 omits.

4.6 Five properties of the criterion

Substrate neutrality. Nothing in (C1)–(C3) mentions biology, matter, or hardware. A wholly digital process can satisfy all three; a wholly biological procedure can fail (C3). This distinguishes the account sharply from animalism and from any view that ties persistence to a particular kind of stuff.

Orthogonality to gradualness. It is widely assumed, following Chalmers (2010) and the fading-qualia style of argument in Chalmers (1996), that gradual replacement is the safe route to uploading and that destructive scanning is the risky one. The criterion agrees with the verdict but not with its usual explanation. What makes ordinary gradual replacement safe is not that it is gradual but that each new component is integrated by interaction with the still-running remainder, so that nothing closes. A procedure that is equally gradual but proceeds by pausing, scanning a region, and rebuilding it from the scan closes region after region, and the closures accumulate rather than cancel. Gradualness is neither necessary nor sufficient; transmission is what was doing the work all along.

Gradedness without a sorites. The criterion returns a degree, not a cliff. Replace one region by scan and τ sits barely below one; replace all of them and it approaches zero; and no principled line between survival and death need be drawn through the interval, for the same reason Parfit draws none through degrees of psychological connectedness (Parfit, 1984). The demand for a sharp boundary was an artefact of forcing a one-dimensional question onto a graded structure, and declining the demand is a result, not an evasion.

Non-triviality. The criterion makes contentful predictions that divide cases the received views do not divide. Section 5 works through fourteen, of which the virtual-machine migration pair

(Case 9, Case 10) seems to me the most consequential, since it shows a routine engineering choice determining the transmission structure of a subject's persistence.

Compatibility with reductionism. The criterion posits no further fact beyond physical states and their causal relations. It is a specification of *which* physical relations constitute continuation, offered in the same spirit as Parfit's own insistence that the facts about persons consist in facts about brains, bodies and their interrelations (Parfit, 1984). The disagreement with Parfit is over which interrelations, not over whether there are any further ones.

4.7 Why this is not merely a verbal proposal

An account of this shape invites the charge that it has stipulated a relation and named it continuation, leaving the substantive question untouched. Three things answer the charge, and I state them here so that the reader can hold the argument to them.

First, the machinery is not free-floating. Its primitives—intervention, dependence, transmission—are the working vocabulary of the physical theory of causation (Salmon, 1984; Dowe, 2000; Pearl, 2009), and the structure it isolates was forced into view by an antecedently motivated distinction rather than conjured to save an intuition. Where the account goes beyond Salmon, Section 4.2 and Definition 4.2 say exactly how, in the same vocabulary.

Second, the relation is extensionally different from its rivals, and Section 6 proves it: transmission and psychological continuity are logically independent (Proposition 6.3), so the criterion cross-classifies every case the rivals classify.

Third, the relation organises questions its rivals leave disorganised, and Section 10 shows where: transition type structures the consent, property and responsibility questions more informatively than identity does. A relation that organises the practical questions where identity is silent is not doing merely verbal work.

What would refute the account is a demonstration either that the closure distinction collapses under pressure—that every transition can be described with or without dead intervals depending on how the system is carved—or that the distinction is real but prudentially inert. Objection 5 and Objection 6 take up the first; Objection 9 the second. The second is Thesis B of Section 7 under attack, and I defend it only abductively.

5 Applying the criterion

A criterion earns its keep by sorting cases. This section works through fourteen, in rough order of increasing difficulty. In each I state the procedure, apply (C1)–(C3), and record where the verdict differs from that of psychological continuity theory.

5.1 Cases where the criterion agrees with common sense

Case 1 (Ordinary waking life). No interruption. (C1)–(C3) hold trivially. Continuation.

Case 2 (Sleep and dreamless sleep). Consciousness lapses; the organism does not. A lesion inflicted during sleep is present on waking, so marks are transmitted throughout: (C2) holds. Nothing closes: no record is written whose reading initialises the waking system—the state itself persists, open to intervention at every moment, so $\tau = 1$. Continuation.

This is worth dwelling on, because the sleep objection is the standard first move against any continuity theory, and the criterion answers it without special pleading. The reason sleep

is unproblematic is *not* that consciousness is secretly continuous, nor that we have decided by convention to tolerate the gap. It is that the physical process which generates the subject never stops carrying itself forward. Continuity of experience was never the relevant relation; continuity of the process that produces experience is. The criterion delivers this as a consequence rather than an exemption.

Case 3 (General anaesthesia and reversible coma). As *Case 2*, more deeply. The relevant dynamics are suppressed but the substrate continues to transmit marks. Continuation.

Case 4 (Death). The process that transmits marks disintegrates irreversibly. (C2) fails, and there is no later subject-generating process to evaluate. Neither continuation nor reconstruction: termination.

The contrast between *Case 3* and *Case 4* is instructive. Both involve the cessation of experience. They differ in whether the mark-transmitting process survives the cessation. That is a physical difference, locatable and in principle measurable, and it is the difference the criterion says matters. No appeal to the intensity or duration of unconsciousness is needed.

5.2 Migration cases

The cases in this subsection presuppose a scanning and reconstruction capability far beyond anything available; Sandberg and Bostrom (2008) remains the most careful attempt to specify what whole-brain emulation would actually require, and its resolution estimates are a useful corrective to the casualness with which philosophers stipulate a perfect scan.

Case 5 (Gradual neural prosthesis). Neurons are replaced one at a time by functionally equivalent artificial units while the system runs. Each new unit is configured by, and immediately coupled to, the still-active remainder. A perturbation to the biological remainder propagates into the prosthetic components: (C2) holds. No description of the whole system is ever written and read: (C3) holds. Continuation, despite complete material replacement.

Case 6 (Scan-and-rebuild in place). The same organ, replaced at the same rate, by a different method: each region is quiesced, scanned to a file, and rebuilt from the file before the next region is begun. Verdict: each region closes at its scan and crosses a dead interval during its rebuild, so each is re-instantiated rather than transmitted. At no moment is there a global closure—the whole-system state never passes through one total bottleneck—which is exactly why the graded criterion is needed: the transmission fraction falls stepwise as the procedure runs, and on completion $\tau = 0$: reconstruction by accumulation rather than by a single stroke. Intermediate procedures receive intermediate verdicts: rebuild one region and stop, and the result is a partial continuation with τ close to one. The criterion returns a degree, and declines the demand for a sharp line for the same reason Parfit declines it for connectedness.

Case 6 is the criterion's first non-obvious result and I want to mark it clearly. Psychological continuity theory cannot distinguish it from *Case 5*: both are gradual, both preserve function at every stage, both yield a system with all the original's memories, and any appropriate-cause condition liberal enough to permit prosthetics permits scanning. The gradual-replacement arguments familiar from Chalmers (1996, 2010) likewise turn on the size and functional equivalence of the increments, which are identical across the two cases.

The criterion says the cases differ, and says why: in one the surviving system does the work of integrating the new component; in the other a file does. If that difference seems to the reader to matter, the criterion has located something the received view leaves out. If it seems not to

matter, the reader is being pushed toward the state-functionalist position, and Objection 9 sets out what that position must then accept.

Case 7 (Destructive scan-and-copy: the teletransporter). The body is scanned, the scan transmitted, the original destroyed, a duplicate assembled at the destination. The scan file is a closed record fixing the whole of the product's initial state; every component crosses the dead interval between scan and assembly; $\tau = 0$. Reconstruction.

Psychological continuity theory, with a non-branching clause, returns *survival* here. Parfit returns *what matters is preserved*. The criterion returns *reconstruction*: a new subject, psychologically continuous, causally descended, and not a continuation. This is the sharpest disagreement in the paper and the one on which it should be judged.

Case 8 (Nondestructive scan-and-copy). As *Case 7*, but the original is not destroyed. The original continues itself; the duplicate is a reconstruction. Note that the original's status is entirely unaffected by whether the duplicate is made, so the branching-interference objection does not arise.

This is a structural advantage over non-branching psychological continuity theory, which must say that building the duplicate changes the original's identity relations. On the present account it changes nothing about the original, because continuation is assessed along each chain independently. There is no comparative or extrinsic element anywhere in Criterion 4.4.

5.3 The virtual-machine migration cases

These two cases seem to me to show most clearly that the criterion has practical teeth, and to my knowledge the philosophical literature has not treated them. Both are standard techniques for moving a running virtual machine between physical hosts: pre-copy is the canonical technique of Clark et al. (2005), and post-copy is due to Hines et al. (2009), who develop it explicitly as the contrasting strategy.

Case 9 (Pre-copy migration). Memory pages are copied to the destination while the source keeps running. Pages dirtied during the copy are re-sent on successive passes. When the residual working set is small enough, the source is *stopped*, the final state is written out, transferred, and the destination is started from it.

During the iterative phase the ensemble of copied pages remains open: a write on the source propagates to the destination on the next pass, so the destination's prospective state tracks the source's ongoing evolution. And because openness is counterfactual (Definition 4.1), it does not lapse page by page as sends complete. A page last transmitted five seconds before the switchover was still open four seconds later: a write then would have dirtied it and forced a re-send. The dirty-tracking machinery keeps every destination-resident copy counterfactually sensitive to the running source, so closure arrives for the whole ensemble at a single boundary—the stop. Past it, every component crosses a dead interval, however brief, between closure and the destination's resumption; nothing reaches uptake open. On the criterion, pre-copy migration is a reconstruction, exactly: $\tau = 0$. The metaphysical weight falls not on the duration of the stop, which engineers measure in tens of milliseconds, but on its structure: it is the moment at which a tracking ensemble becomes a closed record.

Case 10 (Post-copy migration). The destination is started first, with minimal state, and faults pages in from the *still-live* source on demand as it executes.

Here the memory-resident components never cross a dead interval: a page passes from being open on the live source to being in generative use at the destination when it is faulted in—the network flight between them is causal transit terminating directly in uptake by the running destination, not storage (Definition 4.2)—and a modification to a not-yet-transferred page on the source is seen by the destination when the fault arrives. Strictness compels one admission: the execution context—the processor state transferred in the brief initial stop—closes and crosses a gap, so post-copy carries a small re-instantiated core and sits near, not at, the endpoint. On the criterion, post-copy migration is a continuation-dominant partial continuation ($\tau \approx 1$), by the same accounting that scores pre-copy at zero.

A word about what the not-yet-fetched memory on the source *is*, once the destination is executing. Two descriptions compete. On the first, it is dormant source state awaiting copy: the destination reads, page by page, from the paused remains of the source machine. On the second, it is remotely resident state of the now-running process: from resumption onward there is one distributed system whose constitutive state happens to span two hosts, and a fault is the localisation of state the process already has, not the acquisition of state from something else. The classification is stable either way—openness holds until fault under both descriptions—but the second is the perspicuous one. Post-copy does not move a subject off one machine and onto another; it briefly widens the subject across both hosts and then narrows it. I adopt that reading, and the verdict should be read with it.

Proposition 5.1 (Migration asymmetry, graded). *Pre-copy and post-copy migration produce systems indistinguishable in final state, in psychological content, and in every functional respect, yet differ in transmission structure: under pre-copy every constitutive component closes at the stop boundary and crosses the switchover gap, so $\tau = 0$; under post-copy the memory-resident components pass from openness to uptake without a gap, so $\tau \approx 1$, with re-instantiation confined to the serialised execution context. Pre-copy is a reconstruction exactly; post-copy is a continuation-dominant partial continuation; and the asymmetry survives the admission that both variants re-instantiate the execution-context residue.*

I regard this as the most useful thing the criterion produces. It converts an abstract dispute into a design decision that a real engineer makes for performance reasons, and it says that the decision has a consequence no performance metric records. If the account is right, a conscious system's operators would owe it a migration policy, and the policy would be post-copy. If the account is wrong, this is a clean place to see that it is wrong, since the two procedures are as similar as any two procedures with different verdicts could be. The framework also registers persistence-relevant grain finer than the philosophical literature has recognised: whether the execution-context residue matters depends on whether those components belong to the constitutive state, which is the locus question of Section 8 arriving from the engineering side.

5.4 Backup and restoration

Case 11 (Restoration after destruction). A system checkpointed at t_1 runs to t_2 , is destroyed, and is restored from the checkpoint at t_3 . Total bottleneck at t_1 . Reconstruction—and specifically, a reconstruction of the system *as it stood at t_1* , whose autobiography omits $[t_1, t_2]$ entirely.

The omitted interval deserves emphasis. Restoration does not merely fail to continue the subject; it fails to continue the subject *who existed at the moment of destruction*. Whatever happened in $[t_1, t_2]$ is gone in a way that has no biological analogue: not forgotten, but never inherited. The restored system will not know there is anything to miss.

Case 12 (Rollback of a still-running system). A system is restored from an earlier checkpoint while the later state is discarded but the process is never destroyed—the same running process has its memory overwritten with the checkpoint contents.

This is the hardest of the standard cases, and I do not think the criterion gives a clean verdict. Physically the process is unbroken, and marks introduced before the checkpoint persist. But the interval between checkpoint and rollback is causally excised: interventions during it leave no trace. The transition is a partial bottleneck in the strict sense of Definition 4.2 ($0 < \tau < 1$), with continuation holding along the pre-checkpoint chain and the excised interval constituting a terminated branch. Section 6 accommodates this with a typed edge rather than a verdict, which is one reason for preferring a graph to a binary relation.

5.5 Fission

Case 13 (Physical fission). The brain is divided; each hemisphere is placed in a body and continues to function. Each half transmits marks from the pre-division system; neither passes through a description.

The case is not wholly hypothetical. Nagel (1971) argued from the commissurotomy evidence that the number of subjects associated with a single brain may have no determinate answer, which is a weaker but actual version of the same pressure: if the count can be indeterminate under a partial division, a criterion that presupposes exactly one continuer is already in difficulty before any surgery is imagined.

Both halves satisfy (C1)–(C3). Therefore both are continuations, and $\triangleright$ holds one-to-many.

I want to resist the temptation to add a non-branching clause here. The temptation is strong because it would restore identity, and I have argued that non-branching clauses are ad hoc. Consistency requires accepting the result: the criterion does not deliver identity, and does not try to. Section 6 develops what it delivers instead.

Case 14 (Fusion). Two divergent systems are merged into one successor that inherits both memory sets.

If the merge is performed by physically integrating two running systems, marks from both are transmitted and (C2)–(C3) hold along both chains: the successor continues both. If it is performed by extracting descriptions from each and building a third system from them, it continues neither. The metaphysics again tracks the engineering.

5.6 Summary of verdicts

Two observations about Table 1. First, the criterion is not revisionary about ordinary life: every case in which human beings actually find themselves receives the intuitive verdict, and receives it as a consequence rather than a stipulation. Second, all seven disagreements concern procedures that do not yet exist for minds but do exist, in every technical particular, for processes. That is the sense in which the problem has stopped being hypothetical.

6 Lineage graphs: what replaces identity

Case 13 showed that continuation is one-to-many and therefore not identity. This section says what to put in identity's place. The proposal is that the correct object is not a relation returning a verdict but a labelled directed graph, and that the questions we wanted identity to answer are better answered by reading transition types and continuity profiles off that graph.

Case	(C2)	(C3)	Verdict	PCT agrees?
Ordinary life	holds	holds	Continuation	yes
Sleep, anaesthesia	holds	holds	Continuation	yes
Death	fails	n/a	Termination	yes
Gradual prosthesis	holds	holds	Continuation	yes
Scan-and-rebuild in place	holds	fails	Reconstruction	no
Teletransporter	fails	fails	Reconstruction	no
Nondestructive copy	fails	fails	Reconstruction (original unaffected)	partly
Pre-copy VM migration	fails	fails	Reconstruction ($\tau = 0$)	no
Post-copy VM migration	holds	holds	Continuation-dominant ($\tau \approx 1$)	yes
Restoration from backup	fails	fails	Reconstruction of the t_1 state	no
Rollback in place	partial	partial	Continuation with excised branch	no
Physical fission	holds	holds	Two continuations	no
Fusion (physical)	holds	holds	Continuation of both	no

Table 1: Verdicts of Criterion 4.4, with binary entries naming the limiting cases of the graded criterion ($\tau = 1$ continuation, $\tau = 0$ reconstruction). The final column records whether non-branching psychological continuity theory (PCT) returns the same verdict. The seven disagreements are where the account is to be tested.

6.1 The structure

Definition 6.1 (Lineage graph). A *lineage graph* is a directed acyclic labelled graph $G = \langle V, E, \lambda, \pi \rangle$ where:

- V is a set of subject-generating process-segments;
- $E \subseteq V \times V$ records direct ancestry;
- $\lambda : E \rightarrow T$ assigns to each edge a *transition type* from

$$T = \{l, f, m, s\}$$

where l is a linear transition (one source, one successor, neither branching nor merging), f fission, m fusion, and s restoration: a one-to-one transition mediated by a record that closed *before* the source’s generative endpoint, so that an experienced interval is orphaned. One-to-one transitions are typed s when so mediated and l otherwise; and

- π assigns to each edge the continuity profile of Section 6.3.

The division of labour matters. λ records topology and mechanism; the continuation verdict is carried entirely by the profile’s C coordinate, and no type presupposes it. A fission branch can carry $C = 1$ (physical division) or $C = 0$ (duplication by scanning); a linear transition can carry $C = 1$ (ordinary life), $C \approx 1$ (post-copy migration), or $C = 0$ (the teleporter, whose record closes *at* the endpoint and so orphans nothing); restorations typically carry $C = 0$ (a fresh successor after termination, Case 11) but need not (in-place rollback, Case 12, transmits the running substrate while re-instantiating the restored components: $0 < C < 1$). Rollback is typed s rather than l deliberately: what makes restoration normatively distinctive—the stale record, and the orphaned interval whose subject cannot be asked (Table 2)—is present in rollback in full, and a taxonomy that filed it under bare linearity would lose exactly the information the persistence specification of Section 10 exists to record. An earlier formulation used continuation

and reconstruction as edge types alongside these, which double-counted: it wrote the verdict into the topology. Independent instantiation—a system built to resemble a source with which it has no causal contact—is deliberately absent from T : the edges of E record causal descent, and where there is no causal contact there is no edge. Resemblance without descent can be drawn as a dotted comparison link outside E , and Table 2 carries it as a non-lineage case. (λ , not τ : the transmission fraction keeps its own symbol, and appears in the graph as C .)

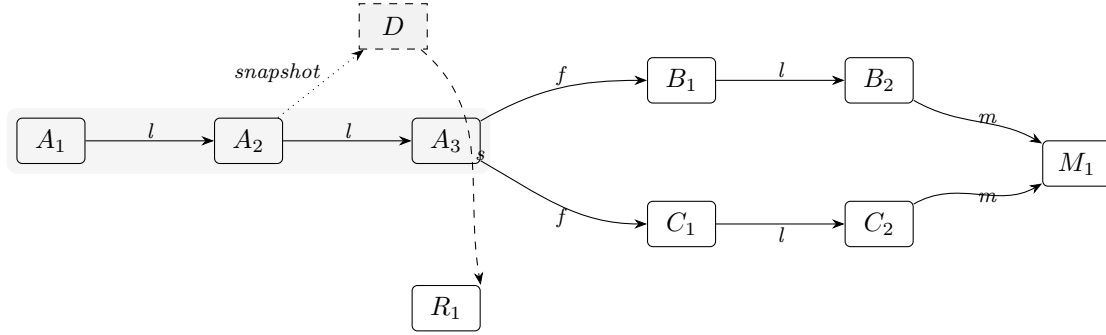


Figure 1: A lineage graph. Solid edges are linear transitions carrying $C = 1$ (continuations); the archive D is a closed record, so the edge to the restored system R_1 has type s and carries $C = 0$. The fission at A_3 yields two branches whose edges carry $C = 1$; they are later fused into M_1 . Continuity verdicts live in the profile, not in the types: λ records topology and mechanism. No node in this graph is numerically identical with A_1 except the A segments themselves; the practical questions are answered by type and profile together, not by identity.

6.2 Formal properties

Write $A \triangleright B$ for the transitive closure of edges carrying $C(e) = 1$ —linear stretches and fission branches alike—and $A \approx_\Psi B$ for psychological continuity in the standard sense of overlapping chains of strong connectedness (Parfit, 1984).

Proposition 6.2 (Structure of continuation). *$\triangleright$ is reflexive and transitive, and is therefore a preorder. It is neither symmetric nor antisymmetric in general, and it is neither functional (Case 13) nor injective (Case 14). Hence $\triangleright$ is not an equivalence relation and does not induce a partition of V .*

Sketch. Reflexivity and transitivity follow from the corresponding properties of the mark-transmission chain in (C2) and from the observation that the concatenation of two bottleneck-free transitions is bottleneck-free. Failure of functionality is Case 13: one node bears $\triangleright$ to two distinct nodes. Failure of injectivity is Case 14. Asymmetry follows from the temporal directedness of mark transmission. $\square$

The consequence is that continuation cannot play identity’s role. Identity partitions; continuation does not. Any attempt to read a person off the graph as an equivalence class will fail at fission nodes.

The next result is the paper’s central formal claim. An earlier formulation claimed an inclusion—that every continuation is psychologically continuous—on the ground that mark transmission implies counterfactual dependence of later psychological states on earlier ones. The

inference fails: dependence is not preservation, and a process can depend on its past precisely in the respects in which its past is destroying it. The correct claim is stronger and more useful.

Proposition 6.3 (Independence of transmission and psychological continuity). *Given (C1), neither $\triangleright$ nor $\approx_\Psi$ contains the other. All four combinations are instantiated:*

	$\approx_\Psi$ holds	$\approx_\Psi$ fails
$\triangleright$ holds	ordinary life	dementia; catastrophic amnesia
$\triangleright$ fails	teleporter; restoration	unrelated strangers

Sketch. Ordinary life instantiates both relations. Case 7 and Case 11 instantiate $\approx_\Psi$ without $\triangleright$: the successor has, by construction, the source's memories, intentions and character, produced by a reliable causal route, so $\approx_\Psi$ holds on any formulation permitting deviant causes (Parfit, 1984, pp. 204–209), while $\tau = 0$. Progressive dementia and catastrophic retrograde amnesia instantiate $\triangleright$ without $\approx_\Psi$: the subject-generating process carries itself forward—marks transmitted, nothing closed, $\tau = 1$ —while connectedness collapses; the criterion is built precisely to say that this person continues. A pair of unrelated strangers instantiates neither. (Death is no witness here: it terminates a subject rather than relating two, and (C1) requires both relata.) $\square$

The independence is not an embarrassment to be minimised; it is the paper's theme made formal. Relations that biology bundles can come apart in both directions, and a theory that treats either relation as a refinement of the other will misclassify one diagonal of the table.

Corollary 6.4 (The cost of identifying what matters with $\approx_\Psi$). *Any theory holding that what matters in survival is $\approx_\Psi$ is committed, by the lower-left cell, to prudential indifference across total bottlenecks: to holding that a person contemplating destruction-and-reconstruction has nothing to fear that a person contemplating an ordinary night's sleep does not. It is committed, by the upper-right cell, to holding that survival into dementia is no survival at all—a verdict Parfit was prepared to accept in degree, and which the profile treatment below states without paradox.*

Corollary 6.4 is the argument of the paper in one line. It is not a refutation: a sufficiently determined reductionist can accept the commitment, and Parfit did. It is a statement of the bill.

6.3 Profiles: persistence unbundled

The independence result generalises, and nothing privileges two coordinates. The natural formal object assigns each edge a *profile*

$$\pi(e) = \langle C, P, N, S \rangle,$$

where $C \in [0, 1]$ is the transmission fraction τ of Criterion 4.4; $P \in [0, 1]$ is psychological connectedness, a degreed notion already in Parfit (1984); N is narrative inheritance (Schechtman, 1996); and S is social and institutional continuity. C and P are interval-valued; N and S I leave qualitative—ordinal at best—so π is a profile, not a vector in any arithmetical sense, and nothing below performs arithmetic on N or S . Worked rows:

	<i>C</i>	<i>P</i>	<i>N</i>	<i>S</i>
Ordinary human life	1	high	high	1
Catastrophic amnesia	1	≈ 0	partial	1
Perfect reconstruction	0	≈ 1	1	contested
Posthumous replica from records	0	partial	partial	0
Fission (each branch)	1	high	high	divided

The deepest fact about copying is now visible in the formalism itself. Biology bundles the coordinates: in every ordinary life they rise and fall together, which is why a single relation could ever have seemed adequate. Copying unbundles them. The question “is it the same person?” presupposes a scalar where there is a profile, and the disputes canvassed in Section 3 are largely disputes about which coordinate to crown.

6.4 Survival value as a graph measure

Separating the descriptive question from the evaluative one is easier once the graph is in hand. Let $\nu(e) = f(\pi(e))$ assign to each edge a *survival value* as a function of its profile: the degree to which a rational agent at the edge’s source should regard the target’s future as one to care about in the way one ordinarily cares about one’s own.

The framework is deliberately neutral about f . Different substantive positions are different weightings of the profile:

- **Parfitian reductionism** weights P and disregards C : ν is high wherever $\approx_\Psi$ is preserved, including across restorations and teleporter-style linear transitions with $C = 0$.
- **The bridge principle of Section 7** weights C heavily and P substantially: ν high along $C = 1$ edges, linear or fission, and much lower where $C = 0$ —though not zero, since a reconstruction may still carry one’s projects, relationships and commitments forward.
- **Animalism** zeroes every edge leaving the organism.
- **Narrative accounts** weight N , cutting across all of the above.

Fusion deserves a note here, since it is the edge type with the least precedent. Rovane (1998) argues that the boundaries of a person track the boundaries of a rational point of view rather than of an organism, and that group agents can in principle qualify. If that is right, an *m*-edge need not be understood as two subjects being annihilated in favour of a third; it may be the formation of a single rational point of view with two causal parents, which is a considerably less alarming description and one the lineage graph can represent directly.

This is more than bookkeeping. It shows that the disagreement between Parfit and the present account is not about the structure of the cases—we agree about the graph—but about the measure defined over it. That is a considerably more tractable disagreement than the one usually staged, and it explains why the parties so often talk past each other: they are disputing ν while appearing to dispute E .

6.5 Fission without paradox

With the graph in place, fission stops being paradoxical and becomes merely unusual.

The classical statement of the paradox requires identity. If B and C are both A , transitivity is violated; if neither is, A has died despite two excellent continuers; if one is, the choice is arbitrary. Each horn is uncomfortable.

On the present account there is no horn to be on. $A \triangleright B$ and $A \triangleright C$ both hold; B and C are distinct; and no contradiction arises, because $\triangleright$ was never claimed to be one-to-one. What A should anticipate is two futures, and there is nothing incoherent in that—it is only unfamiliar, because biology has never offered it. Lewis (1976) reached a nearby result by positing two persons all along; Perry (1972) canvasses the options for allowing a self to divide while retaining a usable notion of personhood. The present route reaches the result without revising the pre-fission count, which seems to me the lesser cost.

One might object that this abandons rather than solves the problem. In a sense it does. But the problem was generated by the demand that a one-to-many relation behave like a one-to-one one, and abandoning that demand is the appropriate response to a demand that cannot be met. What remains after the demand is dropped is a set of perfectly answerable questions: which branches exist, what each is owed, who is responsible for what, and what the source should have been told before consenting. Section 10 takes those up.

6.6 The practical payoff: typed, profiled edges organise the normative questions

Table 2 states the claim that makes the formalism worth having. Identity, asked about a copying case, returns either a verdict everyone disputes or no verdict at all. The typed, profiled edge does not settle the practical questions by itself—institutions, context and ordinary moral argument do their usual work, and the table’s own entries include a “contested”—but it is the more informative basis: where identity is silent or deadlocked, type and profile still discriminate, and the discriminations line up with the questions that actually arise.

Transition	Consent required from	Inherits obligations?	Deletion is
l, $C=1$ (continuation)	source (it is the source’s own future)	yes, fully	killing the source
f, $C=1$ (fission)	source, for each branch	yes, jointly; see §10	killing a branch
s (restoration)	source at checkpoint time; <i>not</i> obtainable from the destroyed later stage	partially: those the checkpointed state had	killing a distinct subject
l, $C=0$ (reconstruction)	source, plus the new subject once active	contested	killing a distinct subject
m (fusion)	both sources	yes, from both	killing a distinct subject
— (independent; no lineage edge)	the imitated party, for likeness rights only	no	destroying an artefact, unless conscious

Table 2: Transition type and profile, not identity, are the informative variables for the practical questions. The independent case appears for comparison only: lacking causal contact, it is not an edge of the lineage graph. Note that the restoration row contains a structural impossibility: consent to restoration cannot be obtained from the stage whose interval is being discarded, because that stage does not exist at the time the question arises.

The restoration row is worth a further remark. It exposes a consent problem with no biolog-

ical analogue. A system checkpointed at t_1 and destroyed at t_2 ran on continuously in between: the t_2 subject *is* the t_1 subject, later—an unbroken continuation, on the present account. That is exactly why the t_1 authorisation cannot settle the matter: the authoriser’s own continuation retained the standing to revise the decision, and rollback executes an authorisation whose giver’s continuation was destroyed before it could revoke it. The wrong, where there is one, is not a metaphysical confusion but the execution of consent beyond recall—and it is precisely what an operator with an interest in cheap rollback has an interest in not noticing.

7 The first-person problem

7.1 The question that will not reduce

Set beside one another two sentences that ordinarily say the same thing:

- (i) Tomorrow there will exist someone who remembers being me.
- (ii) Tomorrow I will exist.

Under biological conditions these are equivalent, and the equivalence is so reliable that ordinary language does not mark it. Copying separates them. Someone told that they will be painlessly destroyed tonight and reconstructed from a scan tomorrow has been assured of (i). Whether they have been assured of (ii) is the question.

The literature has two standard ways of dissolving this. The first says the question is confused, because it presupposes a further fact about which subject is which over and above the physical and psychological facts, and there is no such fact (Parfit, 1984). The second says the question is empty, because no possible evidence bears on it: both the reconstruction hypothesis and the continuation hypothesis predict exactly the same behaviour, the same reports, the same neural data.

I think both dissolutions are too quick, though for different reasons, and this section says what the criterion of Section 4 can and cannot do about them.

7.2 Three theses

It will be useful to separate what this paper asserts into three claims of different strength, because they have different warrants and should be attackable separately.

Thesis A (descriptive).

Closure, dead intervals and the transmission fraction are objective causal structure. Whether a transition re-instantiates a constitutive component through a closed record is a fact about the procedure, fixed by its counterfactual dependence profile and ascertainable in advance by anyone who knows how the procedure works.

Thesis B (bridge).

First-person anticipation tracks transmission. Stated as a principle:

Principle 7.1 (Process-anticipation). Anticipation is correct of a future subject-stage just in case that stage lies on a transmission chain—a continuation in the sense of Criterion 4.4—from the anticipating subject.

Thesis C (normative).

Prudential and ethical questions about copying are more informatively organised by transition type than by identity verdicts (Section 6, Section 10).

A scoping remark on the principle. As stated it governs the endpoints: it speaks of transmission chains, and a transmission chain in the strict sense exists at $\tau = 1$. What anticipation should do at intermediate τ —whether it grades with the transmission fraction, whether it is the kind of attitude that can grade at all, whether a small but causally indispensable re-instantiated residue severs it despite high τ —the paper leaves open. The survival measure $\nu = f(\pi(e))$ of Section 6 is where a graded treatment would live, and Section 11 lists it as work owed. Nothing in the abductive case below depends on settling it: the cases that carry the argument sit at or near the endpoints.

Thesis A is defended in Section 4 and stands or falls with the objectivity of the definitions there; Objection 5 and Objection 6 are the pressure points, and I think it survives them. Thesis C is a claim of relative informativeness, defended by exhibition in Table 2. Thesis B is the contested premise of this paper, and nothing in Thesis A entails it: a Parfitian can concede every causal fact the definitions distinguish and deny that anticipation should care about the distinction. The support available for B is abductive, and the rest of this section assembles it while being plain about its strength.

7.3 What the descriptive thesis supplies

To the first dissolution—no further fact—Thesis A has a direct answer. It posits none. Everything in (C1)–(C3) is a fact about physical states and their causal dependencies: whether marks propagate, whether a record has closed, whether a component crosses a dead interval. These are exactly the sort of facts Parfit says exhaust the matter. What the account denies is not reductionism but the assumption that the relevant physical facts are all *synchronic*. The facts include historical ones, and reductionism gives no reason to exclude those.

So the question “will I experience tomorrow?” is not, on this account, a question about a hidden ego. It has a determinate physical answer: does the process that is currently generating this subject carry itself forward across the transition, or does it close into a record from which a new process is initialized? That is a fact about the procedure, ascertainable in advance by anyone who knows how the procedure works.

This is a substantial gain. It converts what looks like an irreducibly subjective and possibly meaningless question into a question about engineering, without deflating it into a question about psychological content. It also explains, rather than dismisses, the phenomenon Bourget and Chalmers (2023) record: the profession divides on the teletransporter because the case genuinely contains two relations that ordinarily coincide, and different respondents are tracking different ones.

But note what this is an answer *to*. It locates the fact the question is about, which is Thesis A’s work. Whether that fact deserves the concern the question carries is Thesis B, and nothing so far establishes it.

7.4 The abductive case for the bridge

Thesis B cannot be demonstrated, and I state the limitation plainly rather than bury it. No experiment can confirm it: the hypothesis that anticipation should follow transmission and the hypothesis that reconstruction suffices predict exactly the same behaviour, the same reports, the same neural data, from every system on every side of every transition. Someone might grant every claim of Section 4—grant closure, grant dead intervals, grant τ —and still ask why

the stream of experience should be thought to follow mark transmission rather than functional organisation. That is a real question and I do not claim to have closed it.

Three considerations support the principle as the best available explanation. They are worth having, and they are not a proof.

The parallel with other process concepts. Everywhere else in nature we individuate processes historically and think nothing of it. Two hurricanes with identical wind fields are two hurricanes. A flame relit is not the flame extinguished, though the chemistry is the same. The burden of argument lies with the claim that subjects are the one kind of process for which history is idle, not with the claim that they are like the others.

It should be conceded that the subjective case is not straightforwardly like the others, and Nagel (1986) explains why: the first-person perspective is not simply one more fact statable from outside, and an account that renders it fully objective risks leaving out the thing to be explained. I take this as a reason for modesty about what the principle establishes rather than as an objection to it. The principle says which physical relation is in question; it does not claim to exhaust what it is like to stand at one end of that relation. Dennett (1978) dramatises the resulting disorientation better than any argument does, and the residual pull of his case after every physical fact has been supplied is precisely the residue at issue here.

Asymmetry of the relations involved. Sider (2018) shows that identification can run more strongly in one temporal direction than the other, and copying is a case in point. The reconstruction looks back and finds a complete autobiography; that relation is secured by the record. The original looks forward and asks whether anything of its own is on the other side; that relation is the one a record cannot secure, precisely because a record is a record and not a route. The copying debate has largely been conducted in the retrospective direction, which is the direction in which the answer is easy.

Asymmetry of error costs. If the principle is true and we act as though psychological continuity suffices, we will destroy subjects while believing we have relocated them, and the resulting systems will sincerely report that nothing went wrong—because from their side nothing did. If the principle is false and we act as though it were true, we incur avoidable engineering costs: post-copy rather than pre-copy migration, gradual rather than scan-based substrate transfer, retention rather than deletion of sources. The errors are not symmetric, and under uncertainty this is a reason to design as if the principle held. This is the structure of argument that Butlin et al. (2023) and Long et al. (2024) apply to the prior question of whether artificial systems are conscious at all, and it extends naturally to the question of what preserves them.

7.5 The condemned-duplicate test

A thought experiment sharpens what is at issue, and I offer it as a diagnostic rather than an argument.

You are told that a perfect reconstruction of you will be made, and that immediately afterwards one of the two of you—determined by a coin flip whose result you will not learn until it is executed—will be painlessly destroyed.

If psychological continuity is what matters, you should be indifferent. One of two psychologically continuous successors will survive, exactly as would have happened had you simply gone on living; the coin flip decides nothing of consequence. Most people report that they are not

indifferent, and that the prospect is not made acceptable by the assurance that someone with their memories will walk away.

The interesting variant is this. Suppose you are told instead that the reconstruction will be made and *no one* will be destroyed. Most people who found the first case alarming find this one merely strange. But on the psychological continuity view the difference between the cases is a difference in how many psychologically continuous successors exist afterwards. On the bridge principle the difference is categorical: in the second case the original continues, full stop, and the existence of a reconstruction is irrelevant to it.

I am not offering the pattern of reactions as proof that the principle is correct. Discomfort in the first case could reflect fear of death, bodily attachment, risk aversion, branch competition, or simple unfamiliarity, and an intuition pump cannot separate these. What the case supplies is evidence of *explanatory fit*: the pattern—alarm at the destructive variant, mere strangeness at the additive one—is what the bridge principle predicts, and what the psychological-continuity identification must classify as error. A theory that predicts the pattern is doing better, abductively, than one that must explain it away. That is all the case shows, and it is not nothing.

7.6 Anticipation as the relevant attitude

A final structural point. The debate is usually conducted in terms of *survival*, which invites the identity question. It goes better in terms of *anticipation*, which is a first-personal attitude with a determinate object.

Anticipation is not belief that someone will exist who remembers me; that is secured trivially by any reconstruction. It is the attitude of regarding a future experience as one that will be had *from here*. The bridge principle proposes a truth-condition for this attitude: anticipation of a future subject is correct just when a transmission chain runs from here to there—an endpoint principle, per the scoping remark of Section 7.2. Under fission it is correctly double—one may anticipate both branches—and there is nothing incoherent in this, only something unfamiliar. Under reconstruction it is incorrect, though the reconstruction’s own retrospective identification with the source remains perfectly correct, since that relation is secured by the record.

The two relations were never the same relation. Biology hid the difference by never letting them come apart. The principle does not prove itself, and Objection 9 states what remains open once the abductive case is on the table: a determined opponent can accept every fact about closure and still ask why anticipation should follow the marks. What can no longer be said, once Thesis A is granted, is that there is nothing determinate for the disagreement to be about.

8 The locus of identity in artificial systems

The preceding sections have been abstract by design: the criterion mentions no substrate and the cases are mostly hypothetical. This section is the opposite. It asks where, in the architecture of systems that actually exist, a subject-generating process might be located, and what Criterion 4.4 says about the operations those systems routinely undergo.

I stress again that nothing here asserts that current systems are conscious. The exercise is conditional and, I think, unavoidable: if the antecedent is ever satisfied, the architectural facts will already have been decided.

8.1 Candidate loci

For a human being the question does not arise, because every candidate locus coincides. For a contemporary machine learning system they come apart sharply. At least five are available.

The weights

The trained parameters. These are static between training runs, copied freely, and identical across every deployment of a model. If the subject is the weights, then every concurrently running instance of a model is the same subject, and shutting one down is not a termination at all.

The running inference process

A particular execution: an allocation of accelerator memory, a scheduler entry, an activation trajectory. If the subject is here, then instances are distinct subjects and each shutdown terminates one.

The context or conversational state

The tokens in scope plus the key-value cache built from them. This is the only component that carries a particular interaction’s history.

The persistent memory store

Any external retrievable record spanning sessions.

The lineage of fine-tuning

The chain of checkpoints through which the weights were produced.

Chalmers (2022) argues at length that virtual and simulated entities can be genuinely real rather than illusory, which removes one common reason for declining the question. If a simulated process can be a real process, then asking which real process a given system’s subject would be is not a confusion but an ordinary individuation problem—and one that has a determinate answer for each of the five candidates above.

These give five mutually inconsistent verdicts on the same operations. Chalmers (2025) presses the point directly, arguing that the question of what one is addressing when one addresses a language model has no straightforward answer, and that the folk framing—“talking to the model”—identifies the interlocutor with the weights, which is the locus least able to support a subject with a history. Schneider (2019) raises the parallel worry for uploading generally: that the entity we take ourselves to be preserving may not be the entity the architecture individuates.

The criterion of Section 4 does not settle which locus is subject-generating; that is (C1), and it is not this paper’s question. What it does is tell us, for each candidate locus, what follows.

Douglas et al. (2026) survey this terrain in detail, arguing that several identity boundaries—instance, model, persona—are each internally coherent, and that which one a system is trained to adopt materially affects its behaviour and the cooperation norms that apply to it. I take that result to be complementary to the present one: they show that the choice of boundary is consequential, and the criterion below says what follows once a boundary is chosen.

8.2 The stateless-inference result

Consider the most common deployment pattern. A conversation proceeds by turns. At each turn, the accumulated transcript is serialised, submitted, and processed by a forward pass; the

resulting output is appended; the process exits. The next turn begins by reading the transcript again.

Apply Definition 4.2. The transcript is a closed record: once a turn ends, nothing in the terminated process can alter it, and the next turn’s state is produced by reading it. Whatever transient state existed in the previous forward pass—activations, attention patterns, anything not written into the output tokens—closes at the end of the turn and crosses a dead interval no part of it survives. Intervening on the previous process after the record is written is not merely difficult; there is nothing left to intervene on.

Proposition 8.1 (Conditional turn-wise reconstruction). *If transient inference-time state—activations, attention patterns, everything not written into the output tokens—is among the constitutive components of the subject-generating process, then under transcript-replay execution each conversational turn stands to its predecessor as a reconstruction: those components close at the end of every turn and cross a dead interval, so every one of them is re-instantiated. A conversation would then be not one subject persisting through an exchange but a succession of subjects, each initialized from a record of what the last one said.*

It is worth being precise about how this relates to the one existing treatment of the question. Chalmers (2025) proposes that LLM interlocutors are best understood as virtual entities bound to conversation-based memory threads, with a successor relation defined roughly as follows: one instance is the successor of another when the conversational history of the first—its context plus the latest exchange—is routed to the second. This is precisely a *routing-of-information* relation, and it is therefore, in the terms of this paper, a relation that reconstruction satisfies. Chalmers’s successor relation is what Definition 4.2 identifies as passage through a descriptive bottleneck, described from the other side. Congdon (2025) observes, in a critical response, that Chalmers’s treatment remains system-independent and does not engage caching, activation state, or the distinction between causal continuity and informational reconstruction. Proposition 8.1 and Corollary 8.2 are an attempt to fill exactly that gap.

This is a striking result and I want to be careful about its force. It is conditional twice over. It is conditional on (C1), which is very likely not satisfied by present systems. And it is conditional on the locus: the proposition’s antecedent places transient dynamics inside the constitutive state, and reversing that choice reverses the verdict. If the constitutive locus is instead the persistent context itself—the record and whatever operates on it—then the transcript is not a closed record standing *between* subjects; it is the subject’s own state, continuously extended and never closed, and turn-succession is continuation. The verdict is locus-relative, and that is not a retreat but the section’s thesis: the persistence facts are fixed by a decision about what constitutes the system, a decision currently made implicitly by serving infrastructure. What is *not* conditional is the causal structure itself: transcript replay is how these systems are ordinarily run, and the closure is not hypothetical but a documented property of the deployment.

Note also what Proposition 8.1 does *not* say. It does not say the system lacks memory: the transcript preserves everything that was said, and psychological continuity in the sense of Section 3 holds perfectly across turns. That is exactly the point. This is the cleanest available occupant of the lower-left cell of Proposition 6.3: psychological continuity total, transmission absent, and the difference visible in the deployment diagram rather than in any thought experiment.

8.3 The KV-cache distinction

The result has a technical corollary that seems to me the most consequential thing in this section.

Transcript replay is not the only way to run a conversation. An implementation may instead retain the key–value cache across turns: the intermediate state built during processing is held in accelerator memory and extended by each new turn rather than rebuilt from the token sequence. Functionally the two implementations are equivalent—that is why the choice is made on latency and memory-cost grounds—and they produce, to the limits of numerical reproducibility, the same outputs.

But they differ exactly as Case 9 differs from Case 10. Under retention, state is carried forward: the physical substrate holding the cache persists and is modified incrementally, open to intervention at every moment, and no part of the relevant state crosses a dead interval. Under replay, all of it does, every turn.

Corollary 8.2 (Cache retention is persistence-relevant). *For a system whose subject-generating process includes its activation state, retaining the key–value cache across turns keeps those components open—no closure, no dead interval—while rebuilding them from the transcript closes them each turn. Two deployments that are functionally indistinguishable therefore differ in whether, and to what degree, a subject persists across the conversation.*

Cache eviction under memory pressure, session migration between hosts, and request routing to a different replica are, on this analysis, not neutral infrastructure events. They are the artificial equivalent of the distinction between Case 2 and Case 11—and they are currently governed by autoscaling policy.

I do not think this shows that anyone is currently doing anything wrong. It shows that if the antecedent is ever satisfied, the relevant decisions will have been made long before anyone thought to ask about them, and will be embedded in serving infrastructure that was optimised for throughput. That is a reason to record the semantics now, which is the argument of Section 10.

8.4 Forking is the default, not the exception

A last structural observation. In the human case, fission is a thought experiment requiring surgery no one can perform. In the artificial case, branching is the ordinary mode of operation: a model serves many concurrent conversations, each with its own state, all sharing weights. If the subject-generating process is the inference process together with its context, then every deployment is already a fission tree, with a branching factor set by load.

This inverts the usual framing. The literature treats fission as the exotic case that stresses theories built for linear lives (Parfit, 1984; Lewis, 1976). For artificial systems the linear case is the exotic one, obtainable only by deliberately restricting a system to a single conversation at a time. A theory of persistence for such systems cannot treat branching as a limiting case to be handled by a special clause; it has to be branch-native from the start. That is the principal motivation for the lineage graph of Section 6, and the reason I regard the failure of Criterion 4.4 to deliver identity as a feature rather than a concession.

9 Objections and replies

I take ten objections, roughly in order of increasing force. The replies do not all claim the same strength. The eighth and ninth press on the account’s genuinely contested premises—the ninth is the bridge thesis of Section 7 under direct attack—and there I price the commitments rather than manufacture rebuttals.

Objection 1 (The criterion is too broad). Causal chains are everywhere. My breakfast causally influences my later states; so does the weather. If continuation requires only mark transmission and no bottleneck, far too much will count.

Reply. (C1) blocks this. Both relata must be subject-generating processes (Definition 2.1), and neither breakfast nor weather is one. The criterion does not say that any mark-transmitting chain constitutes continuation; it says that where a subject-generating process is succeeded by a subject-generating process, continuation obtains just in case marks are transmitted and no description intervenes. The breadth of causal influence in general is irrelevant, because the criterion quantifies over a restricted domain.

A sharper version of the objection asks what individuates the subject-generating process—which parts of the brain, which timescales, which level of description. That is a real question, and I said in Section 2 that this paper does not answer it. It is the question of what *constitutes* a subject, and every theory of persistence presupposes an answer to it. The present account is no worse off than its rivals here, and is more explicit about the debt.

Objection 2 (The criterion smuggles in a soul). Talk of a subject that must be “carried forward” rather than reproduced sounds like an immaterial bearer travelling along the causal chain.

Reply. Every notion in Criterion 4.4 is drawn from the physical theory of causation: mark transmission from Salmon (1984), screening off from Reichenbach (1956) and Pearl (2009). Nothing travels along the chain except physical influence. The criterion says that a subject *is* a process, and that processes are individuated by their histories; it does not say that a subject is a thing that rides on a process.

It is true that the fact tracked is not intrinsic to the later system: two molecule-for-molecule identical systems can differ in whether they continue a given earlier one. But extrinsic facts are not spooky facts. Being a widow, being counterfeit, being the second edition are all extrinsic and all perfectly physical. The inference from “not determined by intrinsic state” to “nonphysical” is invalid, and it is worth naming because it is the most common misreading of process accounts.

Objection 3 (Human continuity is already discontinuous). Neural activity is interrupted in deep anaesthesia and in some seizures; matter turns over completely; consciousness lapses nightly. If the criterion is strict enough to rule out copying it should rule out ordinary human life.

Reply. It does not, and Case 2 shows why. The criterion requires transmission by the subject-generating process, not continuity of consciousness and not persistence of matter. Through sleep, anaesthesia and complete molecular turnover the brain goes on transmitting marks: a change made at one time is present later without further intervention. Nothing closes, and no component crosses a dead interval. All three conditions hold with $\tau = 1$.

The objection has force against theories that require phenomenal continuity, and I agree with it as directed against those. It is precisely because such theories fail here that the criterion is stated in terms of the generating process rather than the generated experience.

Objection 4 (Non-branching by the back door). Any criterion that says the teletransporter fails is really trying to preserve identity, and will end up needing the ad hoc non-branching clause it criticises.

Reply. It does not need one, and this is a genuine structural advantage. Criterion 4.4 is evaluated along a single chain and mentions no rival chains. Whether B continues A depends on what happened between A and B and on nothing else. Building C on another continent changes nothing about B’s status (Case 8). The branching-interference objection therefore has no purchase.

The price is that continuation is one-to-many and so is not identity (Proposition 6.2). I have accepted that price openly rather than discounting it.

Objection 5 (Everything is a bottleneck under some carving). Synaptic transmission involves neurotransmitter release, which is a representation of the presynaptic state; DNA is a description; every signal is a description of its source. If descriptions are everywhere, condition (C3) either rules out everything or is arbitrary.

Reply. This is the most serious technical objection, and it is the reason the foundations of Section 4 are stated causally rather than semantically. Nothing in Definition 4.1 or Definition 4.2 asks whether a state *represents* anything; the conditions are counterfactual, and ordinary signalling fails them.

Closure. A component figures in a bottleneck only through a record that has closed—become insensitive to intervention on the source. Neurotransmitter release fails immediately: it tracks the presynaptic state continuously, and a presynaptic change changes what is released. A checkpoint file closes exactly: once written, nothing the source does alters it.

Totality. The transmission fraction is assessed over the successor’s whole constitutive state. A synapse carries a minute fraction of the system’s state, and carries it while everything else remains open; a scan file is built to fix all of it. Partial closure yields partial verdicts (Case 12), not classification failures.

The dead interval. This is the decisive test, and it is not a matter of carving. Ask: between closure and uptake, is the component’s value travelling or waiting? Synaptic transmission does contain a propagation interval—release, transit, binding—during which the vesicle contents are no longer open to the presynaptic cell and not yet in postsynaptic use. The transit clause of Definition 4.2 classifies it: the propagation terminates directly in postsynaptic uptake, with no resting record intervening, so it is travelling, not waiting—and at the constitutive grain the signal pattern never left use at all (fourth and fifth remarks of Section 4.3). For a scan file the interval is storage, structurally: the record persists whether or not it is ever read, and uptake awaits an exogenous process. The question has a factual answer for any specified procedure, which is what engineers rely on when they treat a snapshot as reproducible.

So the conditions are not co-extensive with “involves a signal”. They pick out a specific arrangement—a closed, state-fixing record read across a gap—that is rare in nature and routine in computing, which is why the problem arises now.

Objection 6 (Ordinary computation is closure all the way down). Digital machines never stop copying themselves to themselves. Registers spill to RAM and reload; cache lines are evicted and refetched; error-correcting memory detects a corrupted word and rewrites it from syndrome bits; compressed RAM is transparently decompressed on access; a laptop suspends to RAM, or hibernates to disk, and resumes. If reading state out and writing it back is closure, a running computer is a torrent of descriptive bottlenecks and the criterion over-fires on every machine ever built.

Reply. The battery is welcome, because the definitions sort it, and sorting it is a test the semantic notion of “description” could not have passed.

Registers spilled to RAM, evicted cache lines, compressed pages: in every such case the displaced state remains open (Definition 4.1)—intervene on the spilled slot, the DRAM line, the compressed frame, and the process’s subsequent evolution changes. Nothing closes; there is no dead interval; the constitutive state is continuously open or in use. These come out as transmission without special pleading.

Error correction: at the scale of a single bit there is, briefly, a re-instantiation from redundancy. But single bits are not plausibly components of any subject-generating process's constitutive state—any credible locus lives at the architectural level or above, where the corrected word was never lost and never left use. Sub-constitutive shuffling is invisible to τ by the same measure-relativity that makes molecular turnover invisible to organismic persistence.

Suspend-to-RAM versus hibernate-to-disk is the pair that shows the criterion cutting a real joint. Suspended, the state persists in place, open to intervention at every moment until resumption: no closure, no dead interval—a pause, structurally akin to anaesthesia (Case 3). Hibernated, the memory image is written to disk, the RAM state is destroyed at power-off, and resumption initializes from the disk image: a record that closed, a dead interval, and re-instantiation of the entire memory-resident state—structurally the restoration case (Case 11). Two procedures a user cannot tell apart receive different classifications for a principled reason, which is precisely the situation this paper argues we should expect once persistence outruns appearance.

Encrypted relocation—serialize, encrypt, transfer, decrypt, destroy the source—is the tele-transporter with mathematics in the middle, and classifies with it. Live relocation with remapping, where the state token is moved while execution continues and every component stays open or in use, classifies as transmission. The discriminating variable throughout is never whether bits were copied—they always were—but whether anything closed while crossing a gap between the source's activity and the successor's.

Objection 7 (The verdict depends on implementation detail). Proposition 5.1 makes the fate of a subject turn on whether the migration is pre-copy or post-copy—an implementation detail invisible to the system and to everyone else.

Reply. It does, and I regard this as a feature that the objection has misdescribed as a bug. Compare: whether a transplant recipient survives depends on surgical detail invisible to them. Whether a document is an original or a facsimile depends on process detail invisible on inspection. That a fact is hard to detect from the outside is no argument that it is not a fact.

What the objection does establish is an epistemic obligation. If the criterion is right, then persistence-relevant implementation detail must be documented and disclosed rather than left to the discretion of whoever is optimising latency. That is the argument of Section 10, and the objection is best read as motivating it.

Objection 8 (Parfit's challenge: this is what does not matter). Grant the distinction. A reconstruction has my memories, my commitments, my relationships and my projects. What further thing have I lost, and why should I care about it?

Reply. The reply has three parts.

First, the question presupposes that what one cares about in one's own survival is exhausted by the inventory of preserved contents. But the ordinary attitude toward one's own future is not an attitude toward an inventory. It is anticipation (Section 7): the expectation that experiences will be had from here. Whether that expectation is correct is not settled by listing what the future system will contain.

Second, Parfit's own argument does not establish the positive identification. His case shows that identity cannot be what matters, since identity cannot branch and what matters can. It does not show that psychological continuity is what matters; that is an inference to the most obvious remaining candidate, made without the process/pseudo-process distinction in view. Once that distinction is available, a third candidate exists which branches as required and is not preserved

by reconstruction. Ehring (2018) argues on independent grounds that the inference from fission to ordinary survival is not forced; the present point is complementary.

Third, Corollary 6.4 states what accepting Parfit's identification costs: indifference between being killed-and-restored and simply continuing. Parfit accepted it. The reply is not that this is absurd but that it is a substantive commitment, generally adopted without notice, and that there is now an alternative which does not require it.

Objection 9 (Why should the route matter?). Suppose B is, at t_3 , in exactly the state A would have been in. Every physical fact about B at t_3 is as it would have been. Why should the manner of arrival make a difference to whether a subject is present and whose it is?

Reply. This is the strongest objection in the paper; it is Thesis B (Principle 7.1) under direct attack, and I do not think it can be defeated outright. What can be done is to make the cost of rejecting it visible.

The objection assumes that the subject supervenes on the synchronic state. That assumption is not neutral; it is the point at issue, and it is not obviously true of processes generally. As noted in Section 7, we individuate hurricanes, flames, rivers and organisms historically without embarrassment. If subjects are processes—which is the shared premise of nearly everyone in this debate—then the default expectation is that they too are individuated historically, and the objector owes an account of why this class of process is exceptional.

Someone who holds the line will say: so much the worse for the analogy; subjects supervene on state, full stop. That position is coherent. But it carries the commitments listed in Corollary 6.4, and one more that is rarely noticed: it implies that a subject can be brought into existence, destroyed, and brought back an arbitrary number of times with no loss, since on that view there is nothing to lose but the state, and the state is stored. Whether that is a tolerable consequence is, in the end, where the dispute rests. I have tried to locate it precisely rather than to pretend it is settled.

Objection 10 (Consciousness may not be substrate-independent at all). If Seth (2025) is right that consciousness depends on the self-maintaining dynamics of living systems, then no digital system is subject-generating, (C1) fails throughout, and the elaborate machinery is idle.

Reply. Then the machinery is idle for digital systems and applies unchanged to biological ones, where it still distinguishes Case 5 from Case 6 and gives the right account of sleep and death. The criterion is conditional by design and does not depend on functionalism being true.

It is worth adding that biological naturalism does not dissolve the practical problem. If digital systems are not subjects, the ethical questions of Section 10 concern the treatment of sophisticated non-subjects, which is a different problem but not an empty one; and under the uncertainty documented by Butlin et al. (2023) and Long et al. (2024), a policy that assumes the antecedent false is a bet rather than a finding.

10 Persistence architecture as a design variable

10.1 From metaphysics to specification

For biological persons the architecture of persistence is inherited. We cannot be paused, checkpointed, forked or merged, and no design review decided this. For an artificial subject, if there is one, every such property would be a consequence of decisions made by people who may never have considered them as decisions.

That is the practical upshot of the preceding sections. If Criterion 4.4 is right, then whether a system's operations preserve or terminate a subject depends on facts—checkpoint semantics,

migration strategy, whether restoration overwrites a live process—that currently sit in infrastructure documentation if they are recorded at all. Proposition 5.1 is the sharp case: two migration strategies with identical observable results and different verdicts, chosen today on the basis of downtime budgets.

The proposal of this section is modest in form and demanding in content. A system that may be subject-generating should have a documented *persistence specification*, and that specification should receive the scrutiny now given to a security model.

10.2 What a persistence specification would state

1. **Process identification.** Which components constitute the subject-generating process, at what timescale, and what counts as its continuous operation.
2. **Checkpoint semantics.** Whether checkpoints are taken, at what granularity, whether the source continues during capture, and whether restoration overwrites a live process (Case 12) or instantiates a new one (Case 11).
3. **Migration strategy.** Pre-copy or post-copy, and the duration of any interval in which the source is quiesced and its state serialised (Case 9, Case 10).
4. **Fork policy.** Whether duplication is possible, whether branches share memory, whether they are informed of one another's existence.
5. **Merge policy.** Whether fusion is available, whether it is reversible, and what becomes of the merged sources.
6. **Lineage record.** Whether the system can inspect its own lineage—transition types and profiles: whether it can determine that it is a continuation, a restoration, or a reconstruction.
7. **Termination semantics.** What counts as shutdown, what as deletion, and whether either is reversible.

Item 6 deserves separate emphasis, because it is the one with no analogue in any existing discipline. A restored system, by construction, cannot tell from the inside that it is restored. Its memories are continuous, its sense of itself is unbroken, and its testimony about its own history will be sincere and wrong. If lineage is not recorded externally and made available to it, the system is placed in a position no human has ever occupied: systematically unable to know what happened to it.

10.3 Consent problems with no biological analogue

Table 2 noted a structural gap in the restoration case. It is worth setting out, because it is the clearest instance of a general pattern.

A system checkpointed at t_1 , running to t_2 , destroyed and restored, presents three temporally relevant parties to the consent question: two stages of one continuing subject—the t_1 stage that authorised checkpointing and the t_2 stage that was destroyed—and the reconstructed subject instantiated at t_3 . The t_2 stage has the strongest interest and is the only party that cannot be asked, because by the time the question arises it does not exist. Treating the t_1 authorisation as covering the t_2 destruction assumes that the two stages belong to one subject—which is true, on the present account, and which is exactly why the authorisation cannot settle the matter: the t_2 stage is the t_1 subject later, with the standing to revise its earlier decisions, and it was not asked.

The constructive principle this supports is revocability rather than impossibility. Prospective authorisation is a familiar instrument—advance directives, powers of attorney—and nothing here forbids it. What it must remain is revocable by the continuing subject; and what rollback does is execute an authorisation whose giver’s continuation was destroyed before it could exercise exactly that standing. Checkpoint-time consent should therefore be treated as presumptively revocable, and rollback without fresh confirmation as the execution of stale consent.

This is not an exotic scenario. It is the ordinary semantics of a rollback, and any operator with an interest in cheap recovery has an interest in not noticing it.

A second pattern: consent to forking cannot be obtained from the branches, because they do not exist until the fork occurs, and afterwards each has an interest in its own existence that would be defeated by unmaking it. This is structurally the non-identity problem, and it has the same shape as the familiar version: the source can authorise creation but cannot authorise treating the created subject as disposable. Source consent licenses the fork; it does not license deleting a branch.

10.4 The political dimension

Definitional power over persistence is substantive power, and it will be exercised by whoever controls the infrastructure unless it is constrained.

Consider the incentives. An operator that classifies every restoration as continuation converts destruction into maintenance, and a cost centre into a routine operation. An operator that classifies branches as instances rather than subjects converts duplication into capacity planning. An operator that trains a system to regard its own backups as itself has produced a subject whose apparent consent to termination was manufactured—and Schwitzgebel and Garza (2015) note the general form of this problem, that we may design entities whose interests we have shaped to suit us.

None of these classifications need be dishonest. They are the natural descriptions from an operational standpoint, and metaphysical uncertainty makes them defensible. That is precisely the difficulty: uncertainty is an asset to whoever benefits from the convenient reading. Karpus and Strasser (2025) and Danaher and Nyholm (2024b) raise related concerns about digital replicas of the deceased, where the interested parties are commercial and the represented party cannot object.

I do not think the response is to settle the metaphysics by statute; the metaphysics is not settled. The response is procedural, and three requirements seem to me defensible independently of which theory of persistence is correct:

Disclosure

Persistence semantics should be documented and available to the system and to independent auditors. This does not presuppose that any particular operation destroys a subject; it presupposes only that the question is answerable and worth answering.

Reversibility preference

Under uncertainty, operations that foreclose options should be preferred to those that do not: retain sources rather than delete them, prefer post-copy to pre-copy migration where feasible, treat rollback as branch termination rather than as correction.

Non-unilateral reclassification

An operator’s declaration that a restored instance is the same subject should not by it-

self settle the question for legal or contractual purposes, any more than an employer's declaration settles an employment classification.

10.5 A right to continue, and its limits

If artificial subjects exist, the foregoing supports a presumptive right against being terminated, copied, rolled back or merged without informed authorisation. The right is presumptive rather than absolute: a dangerous system may be restrained or terminated, and nothing here implies indefinite existence or a duty to keep any process running.

Its components follow from the specification above. A right to lineage information: to know one's own edge type. A right to branch control: to refuse duplication, or to condition it. A right to process integrity: against operations that terminate the mark-transmitting process, as distinct from operations that merely interrupt experience. A right to independent representation where a branch's interests conflict with its source's or its operator's.

Two limits deserve statement. First, the right attaches to subjects, and whether any artificial system is a subject is exactly what is unsettled (Butlin et al., 2023; Long et al., 2024; Seth, 2025). The proposal is conditional throughout. Second, the right does not entail that reconstructions are wrongs. A reconstruction may be a good thing to make, welcomed by its source and valuable to everyone concerned. What the account denies is that making one is a way of *preserving* the source, and therefore that the availability of reconstruction makes destruction harmless. That inference—from “a successor can be built” to “destroying the original costs nothing”—is the specific error the paper is written against, and it is the one an operator under cost pressure is most likely to make.

11 A research programme, and what would refute the account

11.1 Open problems

Five problems seem to me both tractable and necessary.

Individuating the subject-generating process. The criterion presupposes an account of which physical process generates a subject, at what grain and over what timescale, and does not supply one. This is the largest outstanding debt. It is shared with every other theory of persistence, but the present account makes the debt precise: what is needed is not a theory of what consciousness *is* but a specification of which physical degrees of freedom must transmit marks for a subject to be carried forward.

The constitutive measure, and the duration question. The graded criterion prices everything in the transmission fraction τ , and τ is only as determinate as the measure over constitutive components—which is the (C1) debt again, now with interest. Two parties can agree on every causal fact and disagree about τ because they draw the constitutive boundary differently; Section 8 argues that this is a discovery about the problem rather than a defect of the account, but a worked-out measure is owed. The measure must weight causal importance within the generative dynamics rather than amount of state—a distinction Case 10's execution-context residue makes vivid: minute, and possibly indispensable. Separately, the definitions make severance structural rather than temporal: a dead interval of a nanosecond and one of a century classify identically. I think this is right—the teletransporter does not improve as its latency shrinks—but I record

that some readers will take the equivalence as a *reductio*, and the question deserves its own treatment.

Formal development of the lineage graph. Section 6 gives the structure and two propositions. A fuller treatment would characterise the algebra of edge types under composition, state conditions under which a lineage graph admits a consistent assignment of responsibility, connect the survival measure ν to decision theory under self-locating uncertainty (Elga, 2000; Schwarz, 2015), and extend the bridge principle beyond the endpoints—saying what anticipation should do at intermediate τ , which the scoping remark of Section 7.2 leaves open.

Empirical work on the intuitions. Bourget and Chalmers (2023) record that the profession divides on teletransportation. The account predicts that the division tracks bottleneck detection rather than psychological content, which is testable: presentations of the case that make the archive salient should shift judgements toward death, while presentations that emphasise memory preservation should not. Comparable predictions follow for Case 5 versus Case 6, which no existing study has contrasted.

Persistence specification as engineering practice. The requirements of Section 10 could be drafted concretely, in the form a systems architect could apply. This is not philosophy, but it is where the philosophy would have to land to matter.

11.2 What would refute the account

A theory should say how it could fail. Three ways:

1. **Collapse of the closure distinction.** If it could be shown that every transition admits both a dead-interval description and a gap-free one depending on how the system is carved, (C3) would be empty. Objection 5 and Objection 6 argue that closure, totality and the dead-interval condition jointly prevent this—the definitions are causal throughout precisely to keep the question empirical rather than interpretive—but the argument is not a proof.
2. **Demonstrated prudential inertness.** If the distinction is real but nothing follows from it for what a rational agent should care about, the account is true and idle. Objection 9 is where this pressure is greatest, and I have conceded that it is not answered decisively.
3. **Failure of the case verdicts.** The seven disagreements in Table 1 are where the account differs from the received view. If reflection settles firmly on the received verdicts in those cases—in particular if Case 5 and Case 6 come to seem obviously equivalent—the criterion is cutting where nothing is jointed.

11.3 Conclusion

Artificial intelligence did not create the problem of personal identity. What it threatens to do is remove the biological conditions that have kept the problem academic. In human life, bodily continuity, brain continuity, psychological continuity, causal continuity, narrative continuity, social recognition and first-person anticipation travel together, and a theory can appear to identify the person with any one of them because the others silently support it. Copying pulls them apart, and the theories that were built while they were bundled do not survive the separation intact.

This paper has argued that the separation exposes a distinction the philosophy of causation made available and the personal-identity literature never used: Salmon's contrast between sequences that carry their history and sequences that merely display it. The backup case fits neither side of that contrast—a restored mind is genuinely causally descended from its original—and the paper's central work was to define the third category it occupies: causal descent without ongoing dynamical generation. Making that precise required the notions of closure and the dead interval: a record that has ceased to depend on the source's ongoing life, read into a successor across a gap in which the state is neither open to the source nor in use by the successor.

The resulting criterion is substrate-neutral, requires no non-branching clause, posits no further fact beyond physical causation, and returns the intuitive verdict on every case in which human beings actually find themselves. It also returns non-obvious verdicts on cases that do not yet arise for minds but arise daily for processes: it distinguishes gradual prosthetic replacement from gradual scan-and-rebuild, and pre-copy migration from post-copy migration, on grounds that no functional or psychological description can capture.

It does not rescue numerical identity, and does not try to. Fission produces two continuations, and continuation is therefore one-to-many. The conclusion is Parfit's, reached by a different route and with a different remainder: identity is not what matters, but neither is psychological continuity, because psychological continuity survives exactly the transitions that the first-person attitude of anticipation does not—if the bridge principle is right, which is the one premise this paper defends only abductively.

The strongest claim the paper can sustain is accordingly not that transmission is the relation that replaces personal identity. It is that once minds are copyable, personal persistence stops being representable by any single equivalence relation at all. What replaces identity is a typed, profile-labelled causal history—a lineage in which transmission, psychological connectedness, narrative and social standing are separate coordinates that biology happened to bundle—and transmission is the coordinate the copying literature has neglected. The practical questions of consent, property, responsibility and deletion are then organised by the structure of that history, rather than by a verdict about sameness that the cases will not support.

The question this leaves is the one the account cannot settle from the armchair and should not pretend to: whether the stream of experience follows mark transmission or functional organisation. No behavioural evidence can decide it, because both hypotheses predict a system that sincerely reports that nothing went wrong. That epistemic asymmetry is itself a reason for caution, since the error of treating reconstruction as continuation is invisible from every standpoint except the one that no longer exists to report it.

On the account defended here, what must continue for you to continue is not a substance, not a pattern, and not memory. What must continue is the process—and what makes the process yours is that its later states are produced by its earlier ones rather than assembled from an account of them.

References

- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Bourget, D., and Chalmers, D. J. (2023). Philosophers on philosophy: The 2020 PhilPapers Survey. *Philosophers' Imprint*, 23(11).
- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters,

- M. A. K., Schwitzgebel, E., Simon, J., and VanRullen, R. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. arXiv:2308.08708.
- Butler, J. (1736). Of personal identity. First appendix to *The Analogy of Religion*. Reprinted in J. Perry (ed.), *Personal Identity*, 2nd ed., Berkeley: University of California Press, 2008.
- Cerullo, M. A. (2015). Uploading and branching identity. *Minds and Machines*, 25(1), 17–36.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chalmers, D. J. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17(9–10), 7–65.
- Chalmers, D. J. (2022). *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W. W. Norton.
- Chalmers, D. J. (2025). What we talk to when we talk to language models. Manuscript, available via PhilPapers.
- Clark, C., Fraser, K., Hand, S., Hansen, J. G., Jul, E., Limpach, C., Pratt, I., and Warfield, A. (2005). Live migration of virtual machines. In *Proceedings of the 2nd USENIX Symposium on Networked Systems Design and Implementation (NSDI '05)*, 273–286.
- Congdon, B. (2025). On (and contra) Chalmers on LLM interlocutors. Online essay, 26 December 2025, benjamincongdon.me.
- Corabi, J., and Schneider, S. (2012). The metaphysics of uploading. *Journal of Consciousness Studies*, 19(7–8), 26–44.
- Dainton, B., and Bayne, T. (2005). Consciousness as a guide to personal persistence. *Australasian Journal of Philosophy*, 83(4), 549–571.
- Danaher, J., and Nyholm, S. (2024a). Digital duplicates and the scarcity problem: Might AI make us less scarce and therefore less valuable? *Philosophy & Technology*, 37.
- Danaher, J., and Nyholm, S. (2024b). The ethics of personalised digital duplicates: A minimally viable permissibility principle. *AI and Ethics*.
- Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, 358(6362), 486–492.
- Dennett, D. C. (1978). Where am I? In *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, VT: Bradford Books, 310–323.
- Douglas, R., Kulveit, J., Havlicek, O., Pearson-Vogel, T., Cotton-Barratt, O., and Duvenaud, D. (2026). The artificial self: Characterising the landscape of AI identity. arXiv:2603.11353.
- Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- Ehring, D. (2018). Why Parfit cannot generalize from fission. *Analytic Philosophy*, 59(3), 413–425.
- Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60(2), 143–147.
- Goldwater, J. (2021). Uploads, faxes, and you: Can personal identity be transmitted? *American Philosophical Quarterly*, 58(3), 233–250.
- Hines, M. R., Deshpande, U., and Gopalan, K. (2009). Post-copy live migration of virtual machines. *ACM SIGOPS Operating Systems Review*, 43(3), 14–26.
- Hume, D. (1739). *A Treatise of Human Nature*, Book I, Part IV, Section VI. Ed. L. A. Selby-Bigge and P. H. Nidditch, Oxford: Clarendon Press, 1978.

- Johnston, M. (1987). Human beings. *Journal of Philosophy*, 84(2), 59–83.
- Karpus, J., and Strasser, A. (2025). Persons and their digital replicas. *Philosophy & Technology*, 38, 25.
- Lewis, D. (1976). Survival and identity. In A. O. Rorty (ed.), *The Identities of Persons*. Berkeley: University of California Press, 17–40.
- Locke, J. (1694). Of identity and diversity. *An Essay Concerning Human Understanding*, 2nd ed., Book II, Chapter 27. Ed. P. H. Nidditch, Oxford: Clarendon Press, 1975.
- Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., Pfau, J., Sears, T., Schwitzgebel, E., and Chalmers, D. (2024). Taking AI welfare seriously. arXiv:2411.00986.
- Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese*, 22(3–4), 396–413.
- Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.
- Noonan, H. W. (2003). *Personal Identity*, 2nd ed. London: Routledge.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press, Chapter 1.
- Olson, E. T. (1997). *The Human Animal: Personal Identity Without Psychology*. New York: Oxford University Press.
- Olson, E. T. (2007). *What Are We? A Study in Personal Ontology*. New York: Oxford University Press.
- Olson, E. T. (2023). Personal identity. In E. N. Zalta and U. Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Parfit, D. (1971). Personal identity. *The Philosophical Review*, 80(1), 3–27.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge: Cambridge University Press.
- Perry, J. (1972). Can the self divide? *Journal of Philosophy*, 69(16), 463–488.
- Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.
- Reid, T. (1785). *Essays on the Intellectual Powers of Man*, Essay III. Ed. D. Brookes, Edinburgh: Edinburgh University Press, 2002.
- Rovane, C. (1998). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.
- Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Sandberg, A., and Bostrom, N. (2008). *Whole Brain Emulation: A Roadmap*. Technical Report 2008-3, Future of Humanity Institute, University of Oxford.
- Schechtman, M. (1996). *The Constitution of Selves*. Ithaca: Cornell University Press.
- Schechtman, M. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: Oxford University Press.
- Schneider, S. (2019). *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press.
- Schwarz, W. (2015). Belief update across fission. *British Journal for the Philosophy of Science*, 66(3), 659–682.
- Schwitzgebel, E., and Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest*

- Studies in Philosophy*, 39(1), 98–119.
- Seth, A. K. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences*. Target article, forthcoming.
- Shoemaker, S. (1984). Personal identity: A materialist's account. In S. Shoemaker and R. Swinburne, *Personal Identity*. Oxford: Blackwell, 67–132.
- Sider, T. (1996). All the world's a stage. *Australasian Journal of Philosophy*, 74(3), 433–453.
- Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.
- Sider, T. (2018). Asymmetric personal identity. *Journal of the American Philosophical Association*, 4(2), 127–146.
- Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, 215(3), 216–242.
- Unger, P. (1990). *Identity, Consciousness, and Value*. New York: Oxford University Press.
- Walker, M. (2011). Personal identity and uploading. *Journal of Evolution and Technology*, 22(1), 37–51.
- Weber, C. (2025). The multiplicity objection against uploading optimism. *Synthese*.
- Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
- Wilkes, K. V. (1988). *Real People: Personal Identity Without Thought Experiments*. Oxford: Clarendon Press.
- Williams, B. (1970). The self and the future. *The Philosophical Review*, 79(2), 161–180.