

Dean Zimmerman

*Materialism, Dualism, and "Simple" Theories of Personal Identity*

### **"Complex" and "Simple" Theories of Personal Identity**

Derek Parfit introduced "the Complex View" and "the Simple View" as names for contrasting theories about the nature of personal identity. He detects a "reductionist tradition", typified by Hume and Locke, and continuing in such twentieth-century philosophers as Grice, Ayer, Quinton, Mackie, John Perry, David Lewis, and Parfit himself. According to the Reductionists, "the fact of personal identity over time just consists in the holding of certain other facts. It consists in various kinds of psychological continuity, of memory, character, intention, and the like, which in turn rest upon bodily continuity." The Complex View comprises "[t]he central claims of the reductionist tradition" (Parfit 1982, p. 227).

The Complex View about the nature of personal identity is a forerunner to what he later calls "Reductionism". A Reductionist is anyone who believes

(1) that the fact of a person's identity over time just consists in the holding of certain more particular facts,

and

(2) that these facts can be described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. (Parfit 1984, p. 210)

Take the fact that someone remembers that she, herself, witnessed a certain event at an earlier time.

When described in those terms, it "presupposes" or "explicitly claims" that the same person is involved in both the episode of witnessing and of remembering. Purging the psychological facts of all those that

immediately imply the cross-temporal identity of a person will leave plenty of grist for the mills of psychological theories of persistence conditions. Although veridically remembering may be ruled out, closely related psychological states — and causal connections between them — can be included. As has frequently been noted, Reductionists can replace remembering with a relation of “quasi-remembering” (Parfit 1984, pp. 220-2; Shoemaker 1984, pp. 82-86; see also Shoemaker 1970). Personal identity is not explicitly invoked by describing the experiencing of a certain event by a person at one time, and the occurrence of an *apparent* memory of just such an event by a person later on, which recollection is caused, indirectly, by the original experience in virtue of the normal processes that go on in our brains — whatever those are, exactly. I follow Parfit in calling psychological facts that can be so described “impersonal”. Questions loom about the identity conditions for facts, and about just how fully an impersonal description must characterise a fact in order for that fact to qualify as describable without presupposing or explicitly making claims about personal identity. But I shall not pursue these questions here.

The Complex View favoured by Reductionists is contrasted with the Simple View of an opposing “non-reductionist tradition”. According to Non-Reductionists, ‘personal identity does not just consist in these [psychological and physical] continuities, but is a quite separate “further fact” (Parfit 1982, p. 227; see also Parfit 1984, p. 210). To be a Non-Reductionist, one must “reject either or both of” (Parfit 1984, p. 210) (1) and (2), though they arguably stand or fall together.<sup>1</sup>

---

<sup>1</sup> (2) entails (1), since (2) presupposes the existence of the “more particular facts” in which personal identity is said to consist, according to (1). And, as Marco Dees has pointed out to me, the relation of *partially consisting in* should be transitive and irreflexive; in which case (1) entails (2) as well.

According to Parfit, Non-Reductionists come in two varieties. Some maintain that “[a] person is a separately existing entity, distinct from his brain and body, and his experiences”, such as “a Cartesian Pure Ego, or spiritual substance”, or perhaps “a separately existing *physical* entity, of a kind that is not yet recognised in the theories of contemporary physics”. Other Non-Reductionists accept a “Further Fact View”; they hold that, “though we are not separately existing entities, personal identity *is* a further fact, which does not just consist in physical and/or psychological continuity” (Parfit 1984, p. 210).

The Non-Reductionists include, he says, “the great majority of those who think about the question” (Parfit 1982, p. 227); but this claim stands in need of qualification. He mentions “Butler, Reid, Chisholm, Geach, Swinburne” (Parfit 1982, p. 227); but one wonders who else he has in mind. I should have thought that the dominant philosophical views about persons would all satisfy Parfit’s definition of Reductionism — at least, so long as attention is confined to English-speaking philosophers publishing articles or books since, say, the 1930s. Very few of these philosophers believe in anything like “a Cartesian Pure Ego” or “spiritual substance”. Almost all are materialists, committed to the view that the identity of persons, over time, must be determined by facts about psychological or physical continuities with no need to appeal to souls or any new kind of physical entities “not yet recognised in...contemporary physics”. They may disagree about whether biological or psychological continuities are the basis for the persistence of persons, but they typically attempt to formulate persistence conditions in their favoured terms that leave no room for sameness of persons as a “further fact”.

A closer look at his definition of Reductionism can make one wonder whether *anyone* really qualifies as a Non-Reductionist. Would someone who believed persons are immaterial substances, or a new kind of physical particle, *really* deny Reductionism, given Parfit’s official definition? Well, can an immaterial substance or a new kind of physical particle, and the conditions under which it persists over time, be

described without “presupposing” or “explicitly claiming” anything about the identity of persons? At least some proponents of views along these lines might think so.

Chisholm is one of Parfit’s paradigmatic Non-Reductionists, identified as a defender of the Simple View. The only hypotheses about the nature of human persons that Chisholm took seriously were these two: (i) that each of us is identical with “something of a microscopic nature, and presumably something located within the brain” (Chisholm 1986, p. 75); and (ii) that each of us is “a monad”, an immaterial, indivisible substance (Chisholm 1986, pp. 73, 77). Parfit explicitly consigns both sorts of hypothesis to the category of Non-Reductionism. But it is far from clear that either one is inconsistent with (1) and (2).

By Chisholm’s lights, the proper way for me to determine the nature of my persistence conditions is for me first to try to figure out what kind of thing actually has the psychological properties with which I am directly acquainted. Is it a special kind of physical particle or bit of matter? Then I am that particle or piece of matter, and I have the persistence conditions appropriate to particles or pieces of that sort. Is the bearer of my psychological properties an immaterial substance? Then I am an immaterial substance, with whatever persistence conditions are appropriate to such things.<sup>2</sup> If there are facts — even *brute* facts — about the identity over time of bits of immaterial substance or of some special physical particles that can have psychological properties, then the discovery that *I* am a simple immaterial substance or a special physical particle should lead me to conclude — with Parfit’s Reductionists — that my “identity over time just consists in the holding of more particular facts”, ones that can be described without

---

<sup>2</sup> Chisholm detects the same approach to personal identity in Bayle’s remarks about the rational self-concern of a group of atoms that radically change their form — remarks he cites with approval; see Chisholm 1976, p. 113.

“explicitly” making claims about identity of persons. To further probe this question, I turn to some theses about the relations between persistence and immanent causation.

Chisholm was convinced that, whether we turn out to be physical or immaterial, we must be *fundamental* substances — not mere “modes” of something else, or “*entia successiva*” constituted by different things at different times (Chisholm 1986, pp. 66-71; Chisholm 1976, pp. 104-8). For fundamental substances (ones that are not made out of, or dependent upon, some other kind or kinds of thing), persistence through time plausibly requires the sort of causal dependence of later stages upon earlier stages that is often called immanent causation. Elsewhere, I offer a theory about what such causal dependence should look like. The theory is based upon the idea that the earlier intrinsic features of a thing must be causally relevant to its later intrinsic features (though not in just any old way) (Zimmerman 1997).

Some philosophers — call them “Immanent-Causal Reductionists” — will think that, for fundamental things at least, the facts about persistence consist in nothing more than the facts about immanent-causal connections between stages of the things — i.e., that facts of the first kind are completely determined by, or supervene upon, facts of the second kind. (As Shoemaker points out, one can talk of “stages” in this context while remaining neutral with respect to the doctrine of temporal parts — though many Immanent-Causal Reductionists will believe in temporal parts; Shoemaker 1984, p. 75) The Immanent-Causal Reductionist will have to say that, whenever there are forking paths of immanent-causally connected stages, one must give the same verdict about whether the original thing persists through the apparent fission. There are choices about what the verdict should be: e.g., a fundamental thing ceases to be, replaced by two new ones; or there were two things all along; or a thing becomes multiply located; or.... But a *uniform* answer to all cases of branching paths is required by the Immanent-Causal Reductionist’s commitment to the idea that persistence, for these fundamental things, is fully

determined by the facts about paths of immanent causation. (Similar choices must be made in the case of *converging* immanent-causal paths.)

Others who think immanent-causal continuity is necessary for the persistence of fundamental things may nevertheless suppose that, at least in cases of branching and convergence (and probably elsewhere, too), there are facts about the persistence of fundamental things that are not fully determined by the facts about immanent-causal dependencies. (Chisholm himself would likely have been in this camp.) In a case of branching paths, there are (at least) three genuine possibilities: the original thing ceased to be, or it went to the right while a new thing appeared on the left, or it went to the left while a new thing appeared on the right. (Some will want to allow for a further possibility: the thing became multiply located.) These philosophers posit a further fact about identity, at least for fundamental things, one that is not fully determined by immanent-causal dependencies.

Now, where, within Parfit's taxonomy, should one place an Immanent-Causal Reductionist who believes, with Chisholm, that persons are a special kind of fundamental entity — whether physical or immaterial (however exactly that distinction is to be made out)? Parfit clearly allows that one may make use of psychological terms, when describing the facts in which personal identity is supposed to consist, without violating clause (2) of Reductionism. So the Immanent-Causal Reductionist could say: (a) a person is a thing of any fundamental kind that, at some time or another has psychological properties, or perhaps simply could have psychological properties; and (b) a particular person's persistence conditions just consist in the holding of immanent-causal relations between stages of the kind of fundamental entity he or she (or it) happens to be. This philosopher seems pretty clearly to accept (1) and (2), and therefore to be a Reductionist. But she also seems to satisfy Parfit's example of a paradigmatic Non-Reductionist: she is a believer in a Cartesian Ego or special physical substance, a "separately existing entity".

What should one say about a person who believes that we are fundamental immaterial or physical substances while denying Immanent-Causal Reductionism, maintaining instead that there is a “further fact” about the identity over time of such things? This person seems closer to rejecting Parfit’s (1) (and thereby denying (2)). Still, suppose this philosopher believes in further facts about identity for *all* fundamental things, whether or not they are capable of having psychological properties. She might very well be willing to state very general necessary conditions for the persistence of fundamental things, physical or not — namely, continuity of immanent-causal paths — but then she will say that necessary *and sufficient* conditions will not be forthcoming, because there remains the further question whether the very same fundamental thing is present all along the path. She regards identity over time, for fundamental things, as a kind of “brute fact” above and beyond immanent-causal dependencies. But suppose she is even-handed about the need to posit brute facts, not treating persons as in any way special within the category of fundamental things. She gives very general, though not very informative, persistence conditions for fundamental things of all kinds, ones that do not make use of terms that “presuppose” personal identity or “explicitly” invoke sameness of person over time; and the persistence conditions for persons are subsumed as a special case. Should she really be said to deny (1) and (2)?

Parfit writes as though no version of Reductionism, in his sense, could be consistent with our being “separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events” (Parfit 1984, p. 216). It is not clear to me, however, that Parfit’s arguments against Non-Reductionism must count against views like Chisholm’s, if these theories about personal identity satisfy (1) and (2) — as they would when conjoined with Immanent-Causal Reductionism, for example. The persistence conditions assigned to a Cartesian Ego or a “separately existing” new physical particle by an Immanent-Causal Reductionist may well consist in causal continuities that could, in principle, divide into equally viable streams. If these circumstances require a “no-branching” clause among the facts in which their persistence consists, the resulting views would confront the strongest of Parfit’s arguments for the

conclusion that personal identity is “not what matters” (Parfit 1984, pp. 261-6). They may well be no less (and, for that matter, no more) vulnerable to Parfit’s line of criticism than the more usual forms of Reductionism that he considers.

Be that as it may, Parfit’s more informal characterizations of Reductionism go beyond the idea that personal identity over time does not depend upon further “impersonal” facts; he insists that the facts upon which identity over time depends do not include any “spooky” extra facts about souls or special kinds of matter, but only facts about the mundane stuff our bodies have in common with all sorts of physical objects. The first part of Parfit’s definition of Reductionism seems effectively to be replaced by:

(1\*) that the fact of a person’s identity over time just consists in facts about ordinary material stuff — the kinds of matter our bodies share with non-living, non-sentient things.

I shall understand Reductionism, or the Complex View, as the combination of (1\*) with Parfit’s (2) as a further condition upon the nature of the facts in which personal identity consists — it is needed to rule out a kind of Further Fact View to be discussed below.

I now articulate a reasonably precise thesis that should follow from Reductionism: namely, the supervenience of personal identity upon microphysical facts, or upon microphysical facts supplemented only by “impersonal” psychological facts.

### **Two Doctrines of the Supervenience of Personal Identity Upon the Impersonal**

Philosophers are more reluctant than they used to be to propose that something is “analytic.” Even those who accept an analytic–synthetic distinction will often admit that examples of truly compelling conceptual analyses of interesting notions are thin on the ground. But most of us go on giving philosophical *theories* (of knowledge, right and wrong, objective similarity, causation, etc.), and Parfit is

no exception. Although we might not attribute analyticity to the core tenets of our theories, we do generally take ourselves to be searching for necessary truths that, in some sense, explain the phenomena with which we are concerned. The ethicists (many of them, at any rate) want to know what *makes* something right or wrong, good or bad — what do these normative properties consist in; in virtue of what is an act right or wrong, or a state of affairs good or bad? The search for the grounds for theoretically interesting truths continues in all fields of philosophy, though less frequently under the banner of “conceptual analysis”. Parfit prefers to talk of one fact “just consisting in” other facts. This terminology is, I take it, in the same line of work as talk of the “grounds” of a certain phenomenon; or of what it is “in virtue of which” something is the case.<sup>3</sup>

When a fact is grounded in facts too complicated for us to fully grasp, we may not be able to offer anything like necessary and sufficient conditions for the obtaining of that sort of fact in terms of (what we take to be) its ultimate grounds. We may nevertheless manage to indicate the *kind* of facts upon which we think it must depend. And where there are several rival philosophical theories about some phenomenon, they may nonetheless agree about the *kind* of facts that, in one way or another, ought to provide the grounds for the phenomenon. Parfit’s Reductionism (or, equivalently, the Complex View) is a claim of this sort: whatever the truth is about the conditions that are necessary and sufficient for a person to exist at more than one time, they must ultimately be grounded in facts about ordinary material stuff, and facts that can be described in “impersonal” ways — ways that do not immediately involve the concept of sameness of person.

---

<sup>3</sup> For discussion of various ways to understand a metaphysical “grounding” or “in virtue of” relation, see Fine 1994, 2001 and Schaffer 2009.

When a philosopher claims that facts about the holding of a certain relation just consist in facts about the holding of some other relations, he or she is committed to a “supervenience thesis”: the pattern in which the former relation is exemplified cannot “float free from” the patterns of exemplification of the latter relations; there could not be two worlds just alike with respect to the latter but differing with respect to the former. “Just consisting in” does not, itself, just consist in supervenience. But it entails it. Reductionism, with (1\*) in place of (1), entails the supervenience of personal identity over time upon the exemplification of certain patterns of properties and relations among bits of ordinary matter. I shall take a moment to spell out two supervenience theses in the neighbourhood, one of which will appeal to the physicalistically inclined Reductionist; the other of which is the kind of supervenience that falls most directly out of Reductionism.

I begin with some definitions and simplifying assumptions. Reductionism includes the claim that no “separately existing entity”, in particular no kind of matter “not yet recognised in the theories of contemporary physics”, need be introduced in order to have adequate grounding for facts about persons. New kinds of particles are predicted and discovered all the time, so there is no point trying to make this clause more precise by means of a list. But the atoms in our bodies are made out of particles that can be rearranged to make non-sentient non-persons; and they are subject to a limited number of forces — gravity, electromagnetism, strong and weak nuclear forces — and these forces obey laws that, so far as we know, take no interest in whether the universe contains sentient beings. One kind of Reductionism would be the claim that personal identity over time is fully determined by or grounded in the spatiotemporal relations among the particles of these types (and particles of other kinds, so long as they do not occur only in persons), plus the interactions among these particles that are governed by whatever impersonal forces show up in “final physics”. This would be a kind of Physicalist Reductionism, a view that takes a physicalist gamble: the introduction of consciousness will not require new laws or new forces, let alone new kinds of matter or immaterial souls.

To state a reductionistic claim about the identity over time of particular persons, rather than simply affirming a global kind of physicalism, a Physicalist Reductionist of Parfit's stripe must say: the fact that some person persisted from one time to another just consists in facts about sets of these kinds of particles existing throughout that period, and their being arranged in a certain way and interacting by means of these kinds of fundamentally impersonal forces. It is a thesis about the conditions under which certain batches of particles located at various times *do* constitute a single person existing at all of those times, and the conditions under which they do *not*. The thesis entails a kind of determination of identity by the microphysical: When you have various collections of particles constituting the body of a single person throughout a period of time, there must be impersonal physical phenomena going on involving those microphysical particles and their surroundings which are sufficient to insure that a single person exists during the period.

A canny Physicalist Reductionist will *not* want to say that whether some particles go together to make up a certain person at a time is fully determined by their *intrinsic* interrelations *alone*. Compare a pair of twins, Toothy and Gappy, who are absolutely identical except for the fact that one of them has had the visible parts of all his front teeth removed. (I leave the roots of the teeth intact, so Gappy's gums are not behaving differently from those of Toothy.) The particles making up all of Toothy except for the visible parts of his front teeth may be exactly the same, in their intrinsic arrangement, as the particles making up *all* of Gappy's body. But, in Toothy's case, they do not constitute a human body, only a part of one — whether they make up a whole human body depends upon what relations hold between them and other particles, like the ones nearby in Toothy's front teeth. So, in looking for the underlying physical facts upon which personal identity might depend, one must pay attention not only to relations among the members of a group of particles when asking whether they constitute the same person as someone existing at some other time; one must also pay attention to whether they are embedded in a larger group of particles that could constitute that person *instead*.

I shall help myself to the notion of a certain kind of property of collections of particles: a “complete physical specification” of the “impersonal particles” in our universe — that is, kinds of particles that can exist whether or not there are any sentient beings. Two collections of particles share a complete physical specification if and only if there is an isomorphism between them that preserves all the instantaneous facts about: (i) spatial relations among the members, (ii) all the interactions among them involving basic (impersonal) forces, and also (iii) the kind of physical environment in which they find themselves — for example, proximity to this or that kind of (impersonal) particle, or the presence of basic forces impinging upon them from things not in the collection. The particles making up Gappy’s body do not share a complete physical specification with the set of particles making up Toothy-minus-front-teeth. Although there is an isomorphism that preserves (i) and (ii), it does not preserve (iii).<sup>4</sup> With this notion in hand, it is fairly easy to spell out a supervenience thesis that ought to be part of a Physicalist Reductionism about personal identity over time for creatures with bodies made of our kinds of matter.

- (D1)  $M$  is a Micro-State =<sub>df</sub>  $M$  is a pair consisting of a set of particles and a complete physical specification.
- (D2) The Micro-State  $M$  occurs =<sub>df</sub> the set of particles in  $M$  exemplifies the physical specification in  $M$ .
- (D3)  $H$  is a Micro-History =<sub>df</sub>  $H$  is a series of Micro-States, each one of which is paired with a different positive real number or the number zero.

---

<sup>4</sup> See Zimmerman 1998 for a more detailed presentation of the kind of supervenience thesis sketched here.

- (D4) The Micro-History  $H$  occurs =<sub>df</sub> for every pair of Micro-States  $M$  and  $M^*$  in  $H$ , if  $H$  assigns  $n$  to  $M$  and  $n+m$  to  $M^*$ , then  $M$  occurs  $m$  seconds before  $M^*$  occurs.
- (D5) A Micro-History  $H$  is a continuous part of the history of a single person =<sub>df</sub> (a)  $H$  occurs, (b)  $H$ 's Micro-States form a continuous series ordered by their numbers, and (b) there is a person who, for every time  $t$  at which a Micro-State in  $H$  occurs, is wholly constituted at  $t$  by the set of particles in that Micro-State.
- (D6) Micro-Histories  $H$  and  $H^*$  are indiscernible =<sub>df</sub> (a)  $H$  assigns a Micro-State to some  $n$  if and only if  $H^*$  does; (b) the Micro-States which  $H$  and  $H^*$  associate with any  $n$  have the same complete physical specification; and (c) if there is a particle  $x$  that plays a certain role in the Micro-States that  $H$  associates with  $n$  and  $n^*$ , then there is a particle  $y$  that plays that same role in the Micro-States  $H^*$  associates with  $n$  and  $n^*$ .

Here is a first pass at a statement of the supervenience of personal identity over time upon the microphysical:

- (PR) If a Micro-History is a continuous part of the history of a single person in one possible world, then any indiscernible Micro-History occurring in any possible world is a continuous part of the history of a single person.

Not every materialist will accept (PR); I know that, were I a materialist, I should ask for at least two modifications.

If mental states do not supervene upon the physical, then one might well think that (PR) is false — that personal identity over time does not supervene upon the physical either. But one might still accept Parfit's more general Reductionist thesis: personal identity over time supervenes upon the physical plus the "impersonal" (i.e., personal-identity-neutral) mental states (quasi-memory, etc.) that could be

associated with the history of a single person at a particular time. To add these to the supervenience base, each stage of a history must include facts about whether, in the vicinity of the particles, there is a thinker with impersonal mental states of certain kinds. The specifications of mental states should, like the physical states of the particles, be complete specifications — ways that a mind could be *in toto*, at least with respect to the most fundamental, non-supervenient, impersonal mental properties.

For this non-physicalist Reductionist, a modified supervenience thesis can be constructed:

- (D7)  $M$  is a Micro-Psycho-State =<sub>df</sub>  $M$  is a triple consisting of a set of particles, a complete physical specification, and a complete impersonal psychological specification.
- (D8) The Micro-Psycho-State  $M$  occurs =<sub>df</sub> the set of particles in  $M$  exemplifies the complete physical specification in  $M$ , and also constitute the body of a person who exemplifies the complete psychological specification.

Modifying the rest of the definitions is trivial, and allows for the formulation of a more general supervenience thesis [hence the “G” in “(GR)”], one that even non-physicalist Reductionists could accept:

- (GR) If a Micro-Psycho-History is a continuous part of the history of a single person in one possible world, then, in any world in which an indiscernible Micro-Psycho-History occurs, it, too, is a continuous part of the history of a single person.<sup>5</sup>

---

<sup>5</sup> André Gallois (1998, pp. 248-54) articulates a related doctrine he calls “weak anti-haecceitism”. It is more general than (PR) or (GR) (applying to anything made of parts, not just persons) and stronger in another way too: it implies that if a Micro-History is the history of a

A second reason to doubt (PR) [and (GR) as well] is the fact that nothing in the definition of the occurrence of a Micro(-Psycho)-History requires that any particular causal relationships hold between earlier and later states in the history. Suppose the history of the particles in my body could occur, but with a causal “break” in it — a time at which the occurrence of subsequent Micro-States does not depend, causally, upon the occurrence of earlier ones in the way they *actually* do in *my* history. Arguably, that occurrence of my Micro-History would not be the history of a single person or even of a single organism.<sup>6</sup>

So long as the particles caught up in a Micro-History do not all change at once, the reasonable demand for immanent-causal connections wherever there is identity over time will not allow my history to occur with a *complete* break in causal connections. Still, a further fix-up may well be needed along the following lines:

(PR\*) If a Micro-History is a continuous part of the history of a single person in a possible world  $w$ , then any indiscernible Micro-History occurring in any possible world  $w^*$  is part

---

single person in one possible world, an indiscernible Micro-History that occurs in any world is the history of *that same person*, so long as the same particles play the same roles in each Micro-State. Weak anti-haecceitism is, he thinks, “very plausible”. One might worry that qualitatively similar but numerically distinct objects and events leading up to the beginnings of a Micro-History might make a difference to the identity of the whole made out of the parts caught up in the Micro-History, even if the parts are the same. (PR) and (GR) remain silent on this matter.

<sup>6</sup> Compare Armstrong’s discussion of “immaculate replacements”, Armstrong 1980.

of the history of a single person — so long as, for any  $n$  and  $n^*$ , the Micro-State corresponding to  $n$  is causally dependent (in one way or another) upon the Micro-State corresponding to  $n^*$  in  $w^*$  if and only if they are dependent (in that same way) in  $w$ .

(GR\*) will be understood to be the result of adding the same clause to (GR).

Consideration of abnormal causal mechanisms illustrates the need for the parenthetical “in that same way” to qualify the causal dependence among Micro-States. Suppose time travel is possible, and God exists. If all the atoms in my body at noon today had been caused to jump away from one another exactly then, this organism (and I, if I am this organism) would have ceased to exist. This would be so *even if*, later on, God were to cause these atoms to travel back in time so as to seamlessly continue on the trajectories they were tracing prior to their jump. A *kind* of causal dependence could still hold between the pre-jumping Micro-State of my body and the Micro-States of the same particles, immediately post-time-travel. God notes the state of each particle in my body, just prior to its jump, and then uses that information to send my scattered parts back in time to form a duplicate. God’s activity insures that there is causal dependence between the trajectories of the particles prior to their jump and the trajectories of the time-travelling particles. But the Micro-States in the history of this pseudo-body would not display all the same *kinds* of causal dependence as would the corresponding Micro-States in the otherwise indiscernible Micro-History of my body in the actual world.<sup>7</sup>

---

<sup>7</sup> This would be a sort of “immaculate replacement” without replacement of the fundamental material parts.

## Reasons to Deny (GR\*)

Richard Swinburne accepts “the Simple View” as a good name for his own position — a view he also attributes to Butler, Reid, and Chisholm (Swinburne 1984, pp. 17, 19, and 26). Here is the beginning of Swinburne’s statement of the Simple View (“P<sub>1</sub>” and “P<sub>2</sub>” should be taken to stand for terms that refer to some arbitrarily chosen person or persons — “Obama” and “Nixon”, say, or “Cicero” and “Tully”):

[A]lthough apparent memory and brain continuity are, as they obviously are, evidence of personal identity, they are fallible evidence and personal identity is something distinct from them. Just as the presence of blood stains and fingerprints matching those of a given man are evidence of his earlier presence at the scene of the crime, and the discovery of Roman-looking coins and buildings is evidence that the Romans lived in some region, so the similarity of P<sub>2</sub>’s apparent memory to that of P<sub>1</sub> and his having much the same brain matter, is evidence that P<sub>2</sub> is the same person as P<sub>1</sub>. Yet blood stain and fingerprints are one thing and a man’s earlier presence at the scene of the crime another. His presence at the scene of the crime is not analysable in terms of the later presence of blood stains and fingerprints. The latter is evidence of the former, because you seldom get blood stains and fingerprints at a place matching those of a given man, unless he has been there leaving them around. But it might happen. So the suggestion is, personal identity is distinct from, although evidenced by, similarity of memory and continuity of brain. (Swinburne 1984, p. 19)

Swinburne is surely right in attributing a similar view to Chisholm. Chisholm uses the term “criterion of personal identity” to mean “a statement telling what constitutes evidence of personal identity”, and draws a sharp distinction between the evidence one might have for identifying a person who exists at one time with a person who exists at some other time, and the “truth-conditions” for such claims (Chisholm 1976, pp. 108-13). Both Swinburne and Chisholm include all kinds of physical and

psychological continuities among the sorts of evidence one might have for sameness of person over time, and they insist that personal identity itself is “logically independent” of the holding of such continuities (Chisholm 1976, p. 112; Swinburne clearly agrees; Swinburne 1984, pp. 18-20). Their view is not the triviality that the phrase “is the same person as” means one thing, and “satisfies such-and-such elaborate conditions for being the same person” means something else. Their thesis is much more radical: they reject all philosophical theories that try to state persistence conditions for persons in terms of physical and psychological continuities — at least, physical continuities based upon the behavior of the ordinary matter found in human bodies and brains, and psychological continuities that can be described in an impersonal way.

Swinburne and Chisholm offer thought experiments to support this conclusion — thought experiments that would undermine (GR\*). They describe one or another scenario in terms of the physical and psychological continuities that could hold between persons existing at different times, and then they invite us to *see* that there are two or more genuine possibilities consistent with the events as so far described — possibilities involving differences in which persons exist at which times. There are some possible worlds in which these physical and psychological facts hold, and a single person exists at the beginning and the end of the events described; and there are others, exactly the same with respect to physical and psychological continuities, in which the same person does not make it through the whole episode.

Accepting the failure of (GR\*) on these grounds has drastic consequences: if *any* statement of necessary and sufficient conditions for the persistence of persons over time could be given in terms of physical and psychological continuities, (GR\*)-style supervenience would *have* to hold. So all theories of personal identity that try to formulate such conditions are doomed.

Some philosophers report that, when they conduct these thought experiments, they do not even generate, for them, the *appearance* of more than one possibility. And those who feel the pull of the stories will frequently deny that this appearance is veridical or trustworthy. For the record, I agree with Chisholm's relatively modest claim about the deliverances of these thought experiments: "They seem to me to be worthy of being taken, at least provisionally, as data in our philosophical enquiries about the person; in other words, we should affirm them until we have very good reason for rejecting them." (Chisholm 1970b, p. 188) In this chapter, however, I am not primarily concerned with the status of the thought experiments.<sup>8</sup> I shall instead explore the structure of a certain argument based upon the conclusion that there are two possibilities in these cases. I ask whether the conclusion — the failure of supervenience — can be used to support a robust form of substance dualism. Examination of the argument turns up a relatively clear doctrine that could plausibly be called "Emergent Materialism"; and I point out that Emergent Materialists can make their view consistent with (GR\*) by positing immanent-causal differences due to "emergent properties".

---

<sup>8</sup> In a recent essay, Swinburne defends the deliverances of these sorts of thought experiments. He provides an analysis of the notion of a metaphysically possible world in terms of logically consistent, maximal descriptions couched in terms of what he calls "informative designators"; and he develops a subtle and interesting argument for the conclusion that our failure to see any inconsistency in the thought experiments proves that there are metaphysically possible worlds, in his sense, that are physically and psychologically indistinguishable while differing with respect to facts about personal identity. See Swinburne 2007.

## Two Thought Experiments

Since the stories told by Chisholm and Swinburne have more bite when imagined from one's own perspective, I will use the first person to state them. One thought experiment involves my fission — the division of my body and brain in a way that results in two equally qualified candidates for being me, either of which would understandably be *taken* for me had the other half of me been lost through an injury that simultaneously destroyed all the organs on the other side of my body. Here is a summary of the — by now familiar — story:

FISSION:

I could have been more symmetrical than I am, and could have undergone an operation that brought about my perfect fission. Were that to happen, it would be possible for me to be the person who has the left hemisphere of my brain and other organs from that side; and not to be the person who would be thinking with what had been my right hemisphere. But the reverse is also possible.<sup>9</sup>

There is another line of thought that, although not fully explicit in Chisholm and Swinburne, can be used to support the failure of personal identity to supervene upon physical and psychological continuities. Both philosophers insist that personal identity is a matter of all or nothing, immune to the effects of conventional decisions about where to draw the temporal borders of a thing. But the kinds of physical and psychological continuity that can hold between a person at one time and a person at another time “may be satisfied to varying degrees” (Swinburne 1984, p. 16). When considering just how much of

---

<sup>9</sup> Versions may be found in Chisholm 1969, p. 106, 1970a, and 1976, pp. 111-12; and Swinburne 1973-74 and 1984, pp. 13-20.

one's body or brain could be lost at once, or just how much sudden psychological change could be endured by a person, the proponents of persistence conditions for persons stated in terms of biological or psychological continuity inevitably end up facing borderline cases of the kind of continuity they have selected. The presence of these borderline cases will lead to a failure of supervenience in the following way.

#### FUZZINESS:

Take any kind of mental or physical continuity that could be used in a theory attempting to state necessary and sufficient conditions for the persistence of persons. It will always be possible to describe a series of similar events involving the loss of the relevant kind of continuity, a series that satisfies these constraints: (i) at one end of the series, the event in question is a loss of continuity that seems irrelevant to my surviving; (ii) at the other end of the series, the event is a loss of continuity that seems very likely to be incompatible with my surviving; and (iii) the difference in degree of continuity between each pair of neighbouring cases in the series is always too small for it to be *necessary* that, in the one, I definitely survive, while in the other, I definitely do not. Still, there must always be a definite fact of the matter as to whether I survive any particular episode. So, for at least one of the events in such a series, there are possible worlds in which I survive that degree of loss of continuity, and there are also possible worlds in which I do not survive that very same degree of loss of continuity.

One way to resist FISSION would be to say either: in fission, I would be partially the one, partially the other; or, in fission, I would necessarily die. One way to resist FUZZINESS would be to say: although the differences between some neighboring cases in the loss-of-continuity spectrum may be too small to necessitate that, in one case, I *definitely* survive and, in the next, I *definitely* do not; nevertheless, there will be a transition from cases of continuity sufficient to necessitate definite survival to cases that

merely necessitate its being *less than definitely true* that I survive. Chisholm and Swinburne emphasise the difficulties of these lines of resistance (see, e.g., Chisholm 1976, pp. 110-13; and Swinburne 1984, pp. 13-19).

What the two thought experiments most immediately support is a non-supervenience claim: Where persons go is not *settled*, given all the “impersonal facts” — personal identity could “float free from” such facts, and so (GR\*) is false.

### **Fission and Fuzziness vs. Supervenience Upon the Impersonal**

Fission stories are supposed to undermine both (PR\*) and (GR\*). In the case of perfect symmetry, (PR\*) and (GR\*) require either that the person dies, or goes both ways; but he or she cannot go one way in one possible world, and the other way in the other.

In fuzzy cases, the facts about ordinary matter and impersonal psychological continuities included in micro-histories are imagined to vary smoothly along all the dimensions that could be relevant to sameness of person; but whether something is the body of the same person is (we are supposed to feel) sharp. So there must be worlds in which the same physical and psychological continuities hold, but in one of them the Micro-Psycho-History is the history of a single person, and in the other it is not.

Clearly, accepting the conclusion that (PR\*) and (GR\*) fail due to these possibilities is to treat personal identity over time as involving some kind of “further fact”. But does countenancing these possibilities support the idea that, in addition to the particles that make up the ordinary matter in our bodies, there is also a further substance — be it immaterial or not — that either is the person or is crucial to the ongoing existence of that person?

It has always seemed to me that recognizing these seeming possibilities as genuine *does* support a robust dualism according to which a person is not made entirely of the kinds of matter that appear in

the Micro(-Psycho)-Histories of our bodies. I have accepted what I shall call the “conditional necessity of (GR\*)”: necessarily, if persons were entirely material, (GR\*) would be true. And I am not alone in being attracted to arguments based upon this idea.<sup>10</sup> Here are two, based upon FISSION and FUZZINESS:

#### Fission Argument for Dualism

1. A perfectly symmetrical division of my brain and body could result in two indiscernible Micro-Psycho-Histories which would overlap until the division, and then follow the two halves of my body into different rooms; the histories would be mirror images of one another with respect to causal dependencies among stages; and the procedure would have (at least) the following two genuinely possible (incompatible) outcomes: (a) the left half of my body ceases to be a part of me, and I survive with a body composed entirely of the matter that formerly constituted my right half; and (b) the right half of my body ceases to be a part of me, and I survive with a body entirely constituted by (what was) my left half. (Based on FISSION)
2. If I had no parts other than the particles making up the ordinary matter in this body, then the perfectly symmetrical division resulting in indiscernible Micro-Psycho-Histories

---

<sup>10</sup> Martine Nida-Rümelin makes use of it, appealing just to the case of fission, see Nida-Rümelin 2010, esp. pp. 208-9. She does not regard her support for (her version of) premise 2 as consisting of a “conceivability argument”; but rather as an argument that denial of two genuine possibilities in the case of fission would imply that we (along with any other beings capable of attributing conscious thought to themselves or others) would be radically mistaken about all sorts of important matters.

(with similar causal dependencies) would *not* have the following two genuinely possible (incompatible) outcomes: (a) I survive the loss of my left half, becoming constituted by my right half; and (b) I survive the loss of my right half, becoming constituted by my left half. (From the conditional necessity of (GR\*))

Therefore, I have a part (perhaps an improper part) other than the particles making up the ordinary matter in this body.

#### Fuzziness Argument for Dualism

- 1\*. There is a possible continuation of my actual Micro-Psycho-History in which all the same kinds of causal dependencies hold among stages, but there are two genuinely possible conclusions: (a) I definitely still exist at the end of it, and (b) I definitely do not still exist at the end of it. (Based on FUZZINESS)
- 2\*. If I had no parts other than the particles making up the ordinary matter in this body, then there could be no continuation of my actual Micro-Psycho-History with the same causal dependencies and two genuinely possible conclusions: (a) I definitely still exist at the end of it, and (b) I definitely do not still exist at the end of it. (From the conditional necessity of (GR\*))

Therefore, I have a part (perhaps an improper part) other than the particles making up the ordinary matter in this body.

Most materialists accept a kind of global supervenience of all facts about persons upon the fundamental physical facts about the matter that makes up our bodies; and that sort of supervenience will lead to acceptance of the conditional necessity (and actual truth) of both (PR\*) and (GR\*). But most materialists

are willing to resist these arguments by simply denying their initial premises: FISSION and FUZZINESS have their seductive appeal, but they only *seem* to be possible.

A few materialists take a different approach. They regard physical fission as a genuine possibility for creatures relevantly similar to us; and they believe that, even if both resulting persons were equally similar to the original, there would be a “further fact” about which (if either) was the same as the person who underwent the division. (I find less explicit discussion of FUZZINESS by the materialists I have in mind; but their responses to FISSION suggest analogous things that could be said about FUZZINESS.) So these materialists accept the initial premise of the fission argument. But they deny the conclusion. I will appropriate the over-used, multivalent label “Emergent” as a qualifier for their brand of materialism about persons. The question I want to ask is: Where do the Emergent Materialists think the arguments go wrong?

### **Persons as Entirely Material but “Emergent”**

One natural place for the Emergent Materialist to resist is by giving up (GR\*) (and the conditional necessity of (GR\*)), enabling her to reject premises 2 and 2\* of these arguments. (In the next section, I offer the Emergent Materialist a different strategy: accept (GR\*), and its conditional necessity, but deny the first premises of the arguments from fission and fuzziness.) Let me put some words into the mouth of this sort of Emergent Materialist:

I have none but physical parts, consisting of ordinary types of matter. Nevertheless, at least in cases of fission and fuzziness, one and the same Micro-History or Micro-Psycho-History can be the history of a single person in some possible worlds, but not in others. There is a further fact — whether I go to the left or the right in a fission case, or when exactly I cease to exist in a fuzzy case — that is not settled by all the impersonal facts. Since my comings and goings are not

completely determined by facts about my physical parts, I can rightly be said to be something “over and above” my parts; I show it by my independent spirit. But don’t take the metaphorical phrase seriously. Strictly speaking, there are no such things as spirits (or at least no *human* spirits, if we want to leave God out of this).

Imagine a perfectly symmetrical but otherwise human-like being who undergoes an operation that produces two mutilated but conscious persons waking up in rooms one and two. There is one Micro-Psycho-History consisting of the Micro-Psycho-States of the particles in the original patient and those of the patient in room one; and there is a mirror-image Micro-Psycho-History that begins with the same Micro-Psycho-States but is continued by particles in room two. The Emergent Materialist supposes that, in one possible world, the first Micro-Psycho-History is that of a single person but the second is not; while in another possible world, the reverse is the case. In an operation like this, she says, there is simply a “further fact” about whether the patient loses the right half of his body and moves into room number one, or loses the left half of his body and moves into room number two; and there is a corresponding “further fact” about whether a new person is waking up in room number two or room number one.

Suppose that, as a matter of brute “further fact”, the person recovering in room one is the actual survivor of the operation; and that the patient in room two came into existence upon the operating table. If the Emergent Materialist were to explain the situation to the new man in room two, the latter might well protest. He can make an equally strong case for the claim that *he* was the pre-operative person. But the Emergent Materialist will say:

In many respects, you do seem to be an equally good candidate for being that man. But the fact remains that there was one person who went into the operating room, and he is now waking up

in room number one; and there is no person who went into the operating room and is now in *your* room.

These remarks will be more convincing if the Emergent Materialist can follow them up with the claim that there is *nothing* that went into the operating room looking like a surprisingly symmetrical human being and now lies recuperating in room number two. But an Emergent Materialist who accepts the doctrine of temporal parts cannot say this — at least, not if her metaphysics recognises arbitrary fusions of the temporal parts of persons. When the symmetrically fissioning person shrinks to half his original size and leaves the operating room for room number one, and the mere fission-product wakes up in room number two, this sort of Emergent Materialist must admit that there *is* something that went into the operating room in the form of a whole person, and that is now waking up in room number two. Her claim must be that the only *person* who entered the operating room is now waking up in room number one. The character in room number two will want to know: “Why does my word ‘I’ not refer to a person who existed before the operation? After all, there *is* a thing that existed then, and that was fearfully anticipating the operation, and that is now uttering these words; how could it fail to be a person — in fact, the person I am?”

A temporal parts theorist who denies the existence of arbitrary fusions could give a perfectly good answer to the patient waking up in room two when he asks for a reason to think that he was not the patient who entered the operating room: “Your word ‘I’ doesn’t refer to something with those pre-operative stages as parts because there is *nothing* that has both your current stage and those stages as parts. Those temporal parts and your temporal parts do not form a larger whole.” By contrast, an Emergent Materialist who accepts arbitrary fusions of temporal parts will simply have to regard *being a person* as a property that does not supervene upon the Micro-History or Micro-Psycho-History of an object.

How bad would this be? Worse, I think, than having to accept the failure of microphysical supervenience for the reasons given by Trenton Merricks. Merricks raises difficulties for supposing that the intrinsic properties of collections of particles, plus spatiotemporal and causal relations holding just among the members of the collection, provide a sufficient supervenience basis for facts about whether some particles make up a conscious human being (Merricks 1998a). As Sider has pointed out in his criticisms of Merricks's argument, the materialist is likely to think that *being a human person* is a spatially maximal property, like *being a rock* (Sider 2003, pp. 142-3). A large part of a rock is not a rock — although it would be one, were it not embedded in more rock. Similarly, a large part of a human body, such as a body-minus-its-left-index-finger (to use Merricks's example; Merricks 1998a, p. 62) might be the sort of thing that would *be* a person, were there not more organic matter attached to its hand. Both properties display "border-sensitivity": whether something qualifies as a rock or (given Sider's materialist assumptions) a human person "depends on what is going on around its borders" (Sider 2003, p. 139). Suppose that a materialist wants to regard *being conscious* as intrinsic, and also to follow Sider in treating *being a human person* as a maximal property due to border-sensitivity as in the case of the rock — a human person is the largest of ever so many would-be persons. In that case, she must see a host of conscious non-persons in the vicinity of every human being.<sup>11</sup> Materialists can avoid this conclusion if they follow Sider in accepting the extrinsic nature of consciousness.<sup>12</sup>

---

<sup>11</sup> The moral can be drawn from Merricks 1998a and Olson 1995.

<sup>12</sup> A materialist like Sider can sensibly conclude that consciousness must be extrinsic, in virtue of its analytic ties with *being a person* — something a materialist will naturally take to be border-sensitive. The move is not without its costs, as Merricks points out (2003). I would argue that it is *least* costly for the materialist who regards thinking and experiencing as no more natural or

One might have hoped that an analogous move could be made by the Emergent Materialist who believes in temporal parts and arbitrary fusions; that she could say the failure of supervenience was due to the fact that personhood is an extrinsic matter, and a true supervenience claim can be formulated by appeal to a broader supervenience base — one that is sensitive to the presence of nearby candidates for being a certain person. But the Emergent Materialist's denial that *being a (single) person* supervenes upon Micro-Histories or Micro-Psycho-Histories cannot be rendered benign by claiming that personhood is an extrinsic matter, and only fails to supervene upon Micro(-Psycho)-Histories due to their overly intrinsic nature. A Micro-History includes all the facts about the disposition of matter surrounding the collections of particles caught up in it. So one can tell, just by looking at the Micro-History of the particles that make up the person who survives the operation, that his other half was removed intact and is now recovering nicely in room two.

I conclude, then, that Emergent Materialism and the doctrine of temporal parts make poor bedfellows, barring adoption of mereological principles disallowing arbitrary fusions. Those who reject temporal parts altogether seem to me to be in a dialectically much stronger position. It is not surprising to find

---

fundamental than such obviously non-fundamental activities as executing a karate chop or shaking hands with someone. Arguably, *body-minus-its-left-index-finger* does not do either of the latter things, although it would be doing so were there no left index finger attached to it. Materialists about persons who are property dualists about the phenomenal may not have it so easy. But that is another story.

that all the clearest examples of Emergent Materialists are also opponents of temporal parts — e.g., Trenton Merricks, Tim O'Connor, Jonathan Jacobs, and Lynne Rudder Baker.<sup>13</sup>

### **Holistic Causation to the Rescue?**

In response to the arguments from fission and fuzziness, an Emergent Materialist must either reject (GR\*) (and its conditional necessity), or find some way to deny premises 1 and 1\*. She accepts FISSION (the case upon which I will focus): there is one Micro-Psycho-History that corresponds to the stages in the life of a person who loses half his body and then wakes up in room one, and another that corresponds to this person's pre-operative stages plus the stages of a *different* person who comes into existence on the operating table and wakes up in room two. To accept FISSION and reject premise 1, she must find a causal difference between otherwise exactly similar Micro-Psycho-Histories. Might the causal relations holding between the whole body and just one of the immediately post-operative persons be of a categorically different nature than the ones holding between that body and the other post-operative person? If so, the two Micro-Psycho-Histories might be indiscernible; but they would differ in the kinds of causal dependence holding between stages at the time of the surgery. (GR\*) could then be accepted, but Premise 1 denied.

---

<sup>13</sup> See Merricks 1997, 1998b; O'Connor and Jacobs 2010, pp. 80-2; and Baker 2000, pp. 132-41. For their opposition to temporal parts, see Merricks 1999, O'Connor and Jacobs 2010, p. 74, n. 5; and Baker 2000, p. 22. If I understand Stephen Davis's appeal to the will of God in determining identity (Davis 2010, pp. 26-7), he, too, is committed to the falsehood of (PR\*) and (GR\*). Though I lack hard evidence, I would bet that he, too, rejects the doctrine of temporal parts.

Consider the fissioning symmetrical person, who goes to the left and ends up in room one. In the fission of this organism into two persons, it would be natural to expect that the properties of each resulting person's body, immediately after the operation, would be adequately causally explained by describing two things: the features of just *one half* of the pre-operative person, and the procedure that separated the halves. On this natural assumption, there is no additional property that belongs *only* to the whole pre-operative body immediately prior to surgery, and that causes the post-operative persons to have some intrinsic feature *directly* — i.e., not in virtue of the feature's correlation with contemporaneous intrinsic properties belonging to just one or the other half of the person. I will call this supposition "No Holistic Causation". On this hypothesis, any effect the whole person had upon the intrinsic nature of the person who woke up in room one would be most directly caused by just a part of the pre-operative person. And this does not seem to me to be consistent with supposing that the causal relationship between the whole person and the new guy in room two is of a radically different nature.

Prior to the operation, the left half of the person is not itself a person; and it is causally responsible for the nature of the person in room one. The right half is not a person either; and it is causally responsible for the nature of the person in room two. Considered intrinsically, the two halves generate the two separate persons by an exactly similar process. Assuming No Holistic Causation, the whole person "works through" one of its halves to bring about the existence of a person with certain intrinsic features in room one; he causes something by having a part that causes it. And the whole person also has an exactly similar part — his right half — which, when considered by itself, does exactly the same thing. If something (in this case, a person) uses two exactly similar agents (its two halves) to bring about two exactly similar results by means of exactly similar causal processes, and the thing stands in exactly similar relations to the two agents, then the thing does not cause these two results in significantly different ways.

But relax the No Holistic Causation assumption, and things look altogether different. Suppose that the man is feeling queasy right before the operation; that this feeling of queasiness causes subsequent states of a similar sort; and that it does *not* cause them in virtue of its correlation with more direct causes of subsequent feelings, causes attributable to intrinsic properties of the brain. When the man's pre-operative queasiness causes his post-operative queasiness, it does not do so purely in virtue of intrinsic properties of the half-brain that he kept (since, in general, queasiness is now supposed to have effects that are not entirely due to the physical properties of the man's brain). This opens up space for a difference in the kinds of causal relations holding among Micro-Psycho-States in the two otherwise similar Micro-Psycho-Histories. The pre-operative queasiness of the whole man could directly cause the queasiness felt by the man who is wheeled off to room one; while the similar queasiness presumably felt by the man wheeled off to room two would have to be caused in some other way — perhaps by the physical conditions of the right hemisphere that, upon removal, has become his brain. There would be a kind of causal — or at least counterfactual — dependence holding between the original patient's pre-operative queasiness and that of the new man in room two. If the original patient had not felt queasy, he would not have had the kind of right hemisphere that would tend to produce queasiness in the new man in room two. But, on the holistic causation hypothesis, this would not be the same kind of direct causal dependence that holds between the post-operative queasiness of the man in room one and the pre-operative queasiness of the whole man.

It may well be no coincidence that some of the main defenders of Emergent Materialism (e.g., Merricks 2001, ch. 4, and O'Connor and Jacobs 2010) posit "emergent"<sup>14</sup> properties exemplified by the

---

<sup>14</sup> I should note that Merricks does not himself use the term "emergent" to describe the properties in virtue of which an "object's causal contribution would be...independent of what

person as a whole — properties that are causally efficacious, and not just in virtue of the causal powers exercised by the smaller parts of the body. Indeed, the description of a case of perfectly symmetrical fission given by O'Connor and Jacobs resembles — at least abstractly — the story I told about the whole man's queasiness.<sup>15</sup> Of course, if the holistic causation in question can hold between preoperative patient and *both* fission-products, the Emergent Materialist is once again forced to deny the conditional necessity of (GR\*).

### **Fissioning Simples and Premise 2 of the Fission Argument for Dualism**

The Emergent Materialist asks me — as a defender of the argument for dualism based on FISSION — why I believe that, if something is a mere material object, then in a case of perfectly symmetrical fission, with indiscernible Micro(-Psycho)-Histories, the object *must* go either both ways or neither. The question seems especially pressing when I reflect upon the very different reaction I have to the case of a

---

its atoms were like and so, presumably, independent of what its atoms cause" (Merricks 2001, p. 90); but the label is often applied to such properties by other advocates and opponents of the idea that wholes have properties satisfying this description.

<sup>15</sup> See O'Connor and Jacobs 2010, pp. 80-1. Their "emergent level states" seem to consist mainly of the "particularities" of individuals, however (pp. 74-7), and not phenomenal properties. I am not a fan of the metaphysics of "thin particulars" which they use to explain the notion of a "particularity"; so, were I an Emergent Materialist, I should look for other properties that might have causal powers in some way independent of the activities of a person's proper parts.

simple substance — be it a particle, an intelligent monad, or what-have-you — undergoing apparently symmetrical fission.

#### SIMPLE FISSION:

Imagine a particle,  $x$ , that is buzzing along, minding its own business. Suddenly, it is zapped by some kind of energy, and two particles exactly like  $x$  appear, headed off to the left and right of  $x$ 's original trajectory in what looks like a case of symmetrical fission. Suppose the counterfactuals are the same on both sides: the particle on the left ("Lefty" will be shorthand for this descriptive phrase) would have been spin up had  $x$  been spin up immediately prior to the apparent fission, Lefty would have been moving more rapidly had  $x$  been moving more rapidly then, etc.; and similarly for the one on the right ("Righty", another abbreviated definite description).

What, if anything, is wrong with supposing that there are three genuinely possible endings to SIMPLE FISSION that do not differ in the facts about the particle stages, their environments, and the relations of causal dependence holding between particle stages: (i)  $x$  ceased to be, replaced by two new particles; (ii)  $x$  is Lefty and Righty is a new particle; and (iii)  $x$  is Righty and Lefty is a new particle? Personally, I find myself inclined to say that these do represent three distinct possibilities, each consistent with the case as described. But then why should I have such a different reaction to apparently fissioning composite things, like the Emergent Materialist's physical persons? If I accept these three outcomes as genuine possibilities for SIMPLE FISSION, have I not thereby given up the idea that, when an object is entirely physical, its persistence through time should supervene upon facts that do not directly imply anything about the persistence of the object in question and that concern only its smallest parts and the relations of causal dependence among its stages? If so, have I not lost the right to insist upon the conditional necessity of (PR\*) and (GR\*)?

These supervenience principles describe the way complex wholes depend upon the histories of their parts. When one reaches the level of a partless simple, it is not at all clear that its persistence through time should be thought to consist in anything other than the simple's existing at more than one time. What would a supervenience principle for the persistence of simples look like? It would have to find more fundamental facts that do not directly imply anything about the identities of the kinds of simple in question, and exhibit relations at this level of facts that could plausibly be the grounds of simple-persistence. The friends of temporal parts will have no difficulty finding subjects for such facts, namely, instantaneous temporal parts; but, as shall appear, they should probably resist the idea that SIMPLE FISSION has the three possible endings. Others, who do not regard temporal parts as inevitably accompanying all persistence through time, may well accept that there are three possible endings; but either insist that this is due to the partlessness or fundamentality of the object in question (which distinguishes the case from that of a complex material person), or claim that the story leaves out important details about the kind of causal dependence holding between  $x$  and its offspring, Lefty and Righty.

Those who believe in the doctrine of temporal parts deny that a persisting simple really *is* partless. They are likely to accept the supervenience of the persistence of even the smallest particles upon facts about their instantaneous temporal parts — i.e., to be Immanent-Causal Reductionists about simples. A temporal parts metaphysics may or may not leave room for a deep distinction between immanent causation and other sorts of causal dependence. On a very thin notion of immanent causation, the counterfactuals mentioned in SIMPLE FISSION would settle all the facts about what caused what, and by what means — leaving no room to suppose that the spin and speed of Lefty, say, but not Righty, depended in a special immanent-causal way upon  $x$ 's state at the point of fission. This metaphysician really should accept Immanent-Causal Reductionism, which requires symmetrical treatment of perfect fission; she should deny that the three endings are genuinely possible.

A friend of temporal parts might have a thicker notion of immanent causation; she might think that there is a special kind of causal relation that always holds between stages of the same particle, a relation that is not automatically present whenever there is the kind of counterfactual dependence described in SIMPLE FISSION. She need not immediately draw the conclusion that immanent causation must either fork or stop when *x* is zapped; for all she has been told, the counterfactuals on just one side might be due to immanent causation, the others being the product of another kind of causal relationship. If she thinks that immanent causation is, nevertheless, a relation that *can* hold between the earlier states of one particle and the later states of another, then immanent-causal branching can simply be added to the story. And, again, this temporal parts metaphysician should be an Immanent-Causal Reductionist, not admitting the three possibilities for *x* once *all* the causal facts are stipulated to be symmetrical as well.

Some of us do not believe that persisting through time involves, *as a matter of necessity*, the possession of new, short-lived temporal parts. We have a right to the reaction I reported as my own: SIMPLE FISSION is consistent with each of the three endings. Some of us may deny that any special kind of causal dependence must hold between the way a simple is at one time and the way it was at earlier times — or that any causal dependence must hold at all. These opponents of temporal parts will reject supervenience of the persistence of particles upon other facts. I do not see, however, that doing so should be thought to undermine their confidence in principles like (PR\*) or (GR\*), which pertain to complex wholes.

Other opponents of temporal parts have at their disposal a sort of back-handed way of saving a supervenience principle even for simples. Rejection of temporal parts is naturally coupled with the idea that a robust kind of immanent causation is involved in persistence. When there is genuine persistence of a perfectly natural kind of thing, there is a very distinctive sort of causal relation holding among

stages of the thing — a way for stages in the history of a single particle to depend upon one another that is inevitably accompanied by identity but that stands at a slight conceptual remove from identity nevertheless.

The state or event of a certain fundamental object's having a certain fundamental property at a time can have a tendency, in propitious circumstances, to bring it about that there will subsequently be a fundamental object of the same kind with a fundamental property from the same "determinable" family. But the opponent of temporal parts believes it is possible to have this sort of tendency for two very different reasons: a thing can have it in virtue of a propensity to generate a new thing of the same general type, or it can have it in virtue of a propensity to merely *hang around*. SIMPLE FISSION does not specify whether this second sort of propensity is being exercised at all at the point of fission; nor whether, if it is exercised, the effect occurs towards the right or the left. If this propensity was not triggered by the zapping, *x* ceased to be — however similar to it Lefty and Righty might be, and however robust the counterfactual connections between *x*'s states and theirs.

I find this second response quite plausible: there is a sort of necessarily immanent relation among stages in the history of a fundamental thing, it constitutes a distinguished sort of causation, and it implies the identity of the thing in question. So when I hear the story about *x*'s apparent fission, I see genuinely distinct possible ways in which *x* may have caused the stages of Righty and Lefty; and these differences seem highly relevant to the question whether *x* survived, and, if so, whether it is Righty or Lefty. The supervenience of physical persistence upon the microphysical and the causal, when enriched with this kind of causal dependence, becomes unproblematic. If one supposes that *x* is Lefty, the Micro-History of *x*-up-to-zap-then-Lefty is not exactly like the Micro-History of *x*-up-to-zap-then-Righty; one kind of causal dependence holds between early Lefty stages and *x*'s stage at the zap — immanent causation,

involving the exercise of a propensity to hang around — and some other kind of causal dependence holds between the early Righty stages and  $x$ 's stage at the zap.

So I have found a way to explain why my reaction to the fissioning simple does not betray a lack of commitment to the supervenience of the persistence of physical things upon microphysical histories plus (the right kind of) causal dependencies within those histories. Since the causal dependence is so closely tied to persistence through time, it would not satisfy reductionistic attitudes toward the persistence of simples. But I do not feel much pull toward the idea that identity over time, for simples, must consist in something else.

The differences in direction of immanent causation used in response to SIMPLE FISSION suggest a way for an Emergent Materialist to resist the fission and fuzziness arguments for dualism without *quite* giving up (GR\*) (or its conditional necessity): She can deny premise 1 (and 1\*, though I will focus on FISSION) by positing an important immanent causal difference between otherwise indiscernible Micro-Psycho-States.

In response to SIMPLE FISSION, I gave a sketchy account of a special kind of causation that an opponent of temporal parts might believe in, consisting in the exercise of a propensity to keep existing. Those of us who posit such a causal process may well deny that it occurs in anything other than a fundamental entity. (I should want to leave it a largely empirical question whether subatomic particles persist through time, whether the actual fundamental physical entities are small or large, whether they include fields or a substantival space-time manifold or both. For all I know or care, all fundamental physical entities may consist of series of momentary things, just as the friends of temporal parts allege. I simply see nothing *inevitable* about this.) In order to give the Emergent Materialist a chance, here, I shall suppose that complex organisms can exercise the deeply different kind of hanging-around causation

that I have posited at the level of fundamental persisting things. In particular, I shall allow that it may hold between the stages of an object while that object suddenly loses a large part.

If some state of the pre-operative person immanently-causes some state of the person in room number one, and no state of the man in room two is similarly immanently-caused, premises 1 and 1\* could obviously be denied by the Emergent Materialist. This strategy may make FISSION consistent with the letter of (GR\*); but it is a nice question whether the supervenience principle still captures something of Parfit's Reductionism, once immanent causation is posited as a distinctive kind of causal dependence.<sup>16</sup> On the face of it, appeal to causation that *only* works when an object persists would seem to seriously undermine Reductionism about personal identity. However, if immanent causation can be characterized without any special mention of *persons*, or of psychological properties that imply that a person exists, perhaps the Emergent Materialist will not have violated the spirit of Parfit's Reductionism after all. What is needed at this point is a more precise account than I have given so far of two things: the nature of the "hanging around" kind of causation, and the motivations for advocating Reductionism.

### **Loose Ends and Confessions**

Two of the central topics of this paper are begging for a fuller treatment than I have been able to give them. (1) I am convinced that there is much more to be said in favour of the conditional necessity of (GR\*), and against the Emergent Materialist who would resist arguments for dualism simply by denying (GR\*). (2) I am disappointed by the sketchiness of my attempt to characterise immanent causation, and to distinguish it sharply from other species of causal dependence. I would hope to identify a kind of causal relation definable in terms of the exercise of "strongly identity-entailing conditional powers" — a

---

<sup>16</sup> I am grateful to Marco Dees for posing this nice question.

notion that proved useful when I was grappling with Shoemaker's distinction between "thin" and "thick" properties (Zimmerman 2009, pp. 699-703).

I conclude with a couple of confessions. The conditional necessity of (GR\*) seems to me to glow with the light of Truth, and FISSION and FUZZINESS appeal strongly to my imagination. But if the arguments from fission and fuzziness provided my only reasons for believing in a dualism of persons and bodies, I would regard the case as inconclusive. Conceivability may be an important source of evidence for possibility, but it can lead us astray — especially when used as evidence for the ascription of essential properties to particular individuals, as in the present case.<sup>17</sup>

But other considerations push me towards dualism — or at least away from all the more plausible kinds of materialism, including the versions of Emergent Materialism canvassed here (Zimmerman, 2003, 2004, 2010). In the actual circumstances, then, and despite my qualms about conceivability, the arguments from fission and fuzziness are not irrelevant. Dualism receives a small epistemic boost for me, in virtue of the fact that it respects the conditional necessity of (GR\*) and allows me to trust my modal instincts about FISSION and FUZZINESS.<sup>18</sup>

---

<sup>17</sup> For my criticisms of one argument for dualism that relies upon conceivability in an unsafe way, see Zimmerman 1991.

<sup>18</sup> I learned much from the questions and suggestions of participants in the conference "Diachronic Personal Identity: Complex or Simple?", organised by the Department of Philosophy at the Theological Faculty of the University of Innsbruck, July 2010. I am especially grateful to Richard Swinburne and Hud Hudson. Much later, and in the nick of time, Ofra Magidor showed me that I had somehow begun to use "(GR\*)" to mean two quite different

---

things; a reading group at Rutgers uncovered some technical flaws; and Marco Dees helped me through some confusions. Unfortunately, other comments from these sources, and from audiences at Oxford, Yale, and Buffalo, raised important issues and objections too large to be addressed before the book went to press.

