

The Emergence of Rational Souls

Uwe Meixner

This chapter describes what the emergence of rational souls consists in, explicating the concepts of emergence, soul, and rationality. On the basis of intuitive, yet defended, premises, it argues the existence of rational souls. In a more indirect approach, it points out which basic feature the world must have in order to allow the existence and emergence of rational souls, it gives an argument to the conclusion that the world has this feature in fact, and it finally shows that if the world has the said feature, then the emergence of rational souls is a natural expectation.

1. EMERGENCE

Emergence is, among other things, the coming into existence of something that has never existed before. However, in order to speak of emergence, what emerges does not have to be just new; it must be *ontologically new*: it must be the first exemplar of an *ontological kind* that has never been exemplified before. Thus, my coming into existence at t_2 (my only coming into existence ever) was not an instance of emergence, although I had never existed before t_2 . And the coming into existence at t_1 of the (or of *a*) first human being was not an instance of emergence, either—although this human being had never existed before t_1 , and although the kind *human being* had never been exemplified before t_1 ; for *human being* is not an ontological kind, just a biological one. But the coming into existence at t_0 of the (or of *a*) first living being was indeed an instance of emergence: that living being had never existed before t_0 , and it was the first exemplar of an *ontological kind* that had never been exemplified before t_0 : the (or *a*) first living being was not just new, it was *ontologically new*.

This characterization of emergence implies that in asserting *the emergence of a rational soul*, one is presupposing that *rational soul* is an ontological kind. But that *rational soul* is an ontological kind seems plausible. It even seems plausible that *rational soul* is an *irreducible* ontological kind—which, note, does not already imply that there are rational souls (it merely means: *if* there are rational souls, then they are irreducible entities).

I will call emergence which is such that the ontological kind involved in it is irreducible *radical emergence*. Thus, the emergence of the (or *a*) first rational soul—if it did emerge—was an instance of radical emergence. The emergence of the (or *a*) first living being, however, was very likely not an instance of radical emergence, since *living being* is an ontological kind—indeed—but apparently not an irreducible one. (Today living beings, to the extent that they are just living beings, are almost unanimously regarded as being *purely* physical things—though they were not regarded this way at all times of the philosophical past.)

The preceding remarks on emergence can be summed up and made precise by a sequence of formal definitions:

Suppose *X* comes into existence at time *t*:

- D1 The coming into existence of *X* at *t* is an instance of emergence with respect to *K* if, and only if, (a) *X* never existed before *t* and (b) *K* is an ontological kind that was never exemplified before *t*, and *X* exemplifies *K* at *t*.
- D2 The coming into existence of *X* at *t* is an instance of radical emergence with respect to *K* if, and only if, it is an instance of emergence with respect to *K*, and *K* is an irreducible ontological kind.
- D3 The coming into existence of *X* at *t* is an instance of emergence (simpliciter) if, and only if, it is true for some *K* that the coming into existence of *X* at *t* is an instance of emergence with respect to *K*.
- D4 The coming into existence of *X* at *t* is an instance of radical emergence (simpliciter) if, and only if, it is true for some *K* that the coming into existence of *X* at *t* is an instance of radical emergence with respect to *K*.

If the coming into existence of *X* at *t* is an instance of emergence with respect to *K* and if there is no coming into existence of anything else at *t* that is an instance of emergence with respect to *K*, then *X* can be described as *the first* [exemplar of the ontological kind] *K*, and the coming into existence of *X* at *t* is a case of *singular emergence* with respect to *K*. If besides the coming into existence of *X* at *t* also the coming into existence of something else at *t* is an instance of emergence with respect to *K*, then *X* can still be described as *a first* [exemplar of the ontological kind] *K*; but the coming into existence of *X* at *t* is not a case of singular emergence with respect to *K*: it is part of a case of *plural emergence* with respect to *K*. Since plural emergence with respect to *K* requires *exact simultaneity* of the several (at least two) comings-into-existence participating in it, plural emergence with respect to *K* is bound to be a very rare thing, no matter which ontological kind *K* we are looking at. Perhaps there are no (true) cases of plural emergence at all. Nevertheless, plural emergence (like causal overdetermination) can certainly not be excluded on a priori grounds.

According to the preceding definitions, emergence is—relatively or simpliciter—a property of *comings-into-existence*. But, not infrequently,

emergence is also ascribed to individuals that are not event-like entities: there are assertions in singular of the form “the first NP emerged” (where “NP” stands for a noun-phrase)—for example: “The first living being emerged”—and there are assertions in plural of the form “NPs emerged”—for example: “Living beings emerged”. Such assertions are *derivative, analogical* ways of speaking.

“The first NP emerged” just means the same as “There is a time t and an X which is such that X came into existence at t , and the coming into existence of X at t is an instance of emergence with respect to NP [where NP is the *kind* that corresponds to the noun-phrase NP], and for every Y that came into existence at t and that is different from X : the coming into existence of Y at t is *not* an instance of emergence with respect to NP ”.

And “NPs emerged”, in one interpretation (the *lectio faciliior*), just means the same as “There is an X and a time t such that X came into existence at t , and the coming into existence of X at t is an instance of emergence with respect to [the kind] NP ”. Accordingly, the plural in the statement “NPs emerged” is not to be taken referentially seriously: one might as well have said, “A [or *some*] NP emerged”. Note that, according to the present interpretation, “NPs emerged” is logically equivalent not only to “A NP emerged” but also to “[The kind] NP emerged”. And note that *this* interpretation is precisely the one in which the statement “Rational souls emerged” *underlies* the title of this chapter, “The Emergence of Rational Souls”. Hence I might logically equivalently (and without ambiguity) have called this chapter “The Emergence of a Rational Soul” or “The Emergence of *Rational Soul* [the kind]”. But these alternative titles would not have been as catchy as its present title.

In another interpretation of “NPs emerged” (the *lectio difficilior*), the phrase means the same as “There is a time t and an X such that X came into existence at t , and the coming into existence of X at t is an instance of emergence with respect to NP , and there is a Y different from X that also came into existence at t and the coming into existence of Y at t is also an instance of emergence with respect to NP ”. In this interpretation the plural in “NPs emerged” is taken seriously, and “NPs emerged” is logically equivalent to “Several NPs emerged”.

Note finally that statements of the following forms are pairwise logically equivalent:

The first NP emerged.	NP emerged singularly.
NPs emerged ₁ [<i>lectio faciliior</i>].	NP emerged.
NPs emerged ₂ [<i>lectio difficilior</i>].	NP emerged plurally.

2. SOULS

Presuming that *rational soul* is an irreducible ontological kind, what are the characteristics of that kind? In answering this question, I shall first look at

the concept of *soul* in and by itself, leaving *rationality*, for the moment, out of consideration. The notion of soul is a very old one, with a very long tradition in the history of ideas. Given this tradition, very many philosophical, religious, theological, psychological, and folk-psychological ideas, perspectives, and connotations have accumulated around that notion. I herewith present the definition of a sober, strictly philosophical concept of soul; it is inspired by the tradition, but at the same time highly selective with regard to the riches of that tradition:

- D5 *X* is at *t* a soul of *Y* if, and only if, (1) *Y* is an animal that is conscious at *t*, (2) *X* is a non-physical individual without inherent temporal dimension, (3) *X* is a subject of the consciousness of *Y* at *t*, and (4) if *Y* is up and about in the time-stretch following *t*, then *X* causes some movement¹ of *Y* after *t* in the light of the consciousness of *Y* at *t*.

According to D5, the concept of *soul*—or better: *soul-of*—is a relational concept, a rather special form of time-dependent *being-of* (other, more commonplace forms of time-dependent *being-of* are wife-of, home-of, father-of, son-of, etc.). Note also that according to D5 the concept of *soul* is identified with the concept of *active soul* (see condition (4)); this should be kept in mind.

There is a *normal way* of *X* being at a time a soul of *Y* (but D5 does not propose a priori that it is the *only way*). This is the normal way:

The normal time-dependent <animal, soul>-relationship

- (a) For some time *t*: *Y* is an animal that is conscious at *t*.
- (b) For every time *t*: if *Y* is an animal that is conscious at *t*, then *X* is at *t* a soul of *Y*, and nothing else but *X* is at *t* a soul of *Y*.
- (c) There is no *Y'* distinct from *Y* such that at any time *t* *X* is a soul of *Y'* at *t*.

If *X* and *Y* fulfill the condition of normalcy that is specified by the conjunction of these three statements, then one can address *X* as *the soul of Y* (this is based on (a) and (b)) and as *the soul of Y alone* (this is based on (a), (b), and (c)). And vice versa: if one can address *X* as *the soul of Y* and also as *the soul of Y alone*, then *X* and *Y* fulfill the condition of normalcy that is specified by the conjunction of those three statements. (I leave it to the reader to consider which *non-normal* animal-soul-relationships are not excluded by D5, and which of those may plausibly be assumed to actually occur in the light of psychopathological evidence.) Now, the simple sense in which the word “soul” is meant in the title of this essay is this:

- D6 *X* is a soul (*simpliciter*) if, and only if, there is an animal *Y* such that *X* is *the soul of Y*.

If X is at time t a soul of Y , then there is a connection between X and Y in virtue of which X is of Y at t . What is the nature of that connection? Generally speaking, it is a connection between a non-physical individual and a *physical one*; this follows on the basis of D5 and the fact that *being an animal* entails *being a physical individual*. In part, the connection in question is effected by *consciousness*, and in part, it is effected by *causal power*. The two aspects of the connection are closely related to each other—not only on the basis of the obvious fact that each of the two connects the same thing, X , with Y , but also on the basis of a functional relationship between them (indicated in D5 by the words “causes [. . .] in the light of the consciousness”).

I shall first consider the consciousness-related part of the connection of X with Y . If Y is an animal that is conscious at t , then there is such a thing as *the consciousness of Y at t* , in other words: *the totality* of those experiences that have both the index Y and the index t . The index Y becomes (not immediately, but *mediately*) attached to that totality by the fact that it is the brain of Y that produces the experiences in the totality (and forms them into a homogeneous whole); the index t becomes attached to that totality by the fact that the experiences in the totality all occur at time t (for this, one need not necessarily assume that t has no extension). Thus, *the consciousness of Y at t* is the totality of all experiences produced by the brain of Y that occur at t . But, *note*, this does not make Y or the brain of Y a *subject of the consciousness of Y at t* . Rather, such a subject is itself a product of the brain of Y , which is embedded in, or better: *implied by*, the (brain-produced) consciousness of Y at the given time t .

That a subject of the consciousness of Y at t is neither Y nor the brain of Y is confirmed by elementary phenomenology (even *without* reference to evidence that shows beyond reasonable doubt that a subject of the consciousness of Y at t is a *dependent product* of the brain of Y , and hence cannot be identical with either Y or its brain): I, for example, am not merely a , but *the subject of the present consciousness of this animal*, UM; but I am neither UM nor its brain. Why? Because there is at present, that is: at t^* , a certain, narrowly circumscribed location L in space from which I see the world. *If I am located anywhere in space at t^* , then I am located in L at t^** (*this* is what phenomenology rather convincingly tells me; for now, at t^* , my right toe is three feet, in this particular direction, away from L —and away from *me*, it seems). Suppose now that I am located somewhere in space at t^* ; hence, according to the conditional just asserted—the *conditional in italics*: CI—, I am located in L at t^* , and hence I am neither this animal nor its brain (because neither of the two is located in L at t^*). Suppose then, on the contrary, that I am not located anywhere in space at t^* ; hence, again, I am neither this animal nor its brain (because both are located somewhere in space at t^*).

In fact, if CI is true (and I believe it is true), then I am distinct from every physical individual. This is seen by the following argument: Since I exist at t^* , [1] I am distinct from every physical individual that does not exist at t^* .

And [2] if I am nowhere in space at t^* , then I am distinct from every physical individual that exists at t^* (because every such individual is somewhere in space at t^*). Finally, [3] if I am somewhere in space at t^* , then I am *in L* at t^* (according to *CI*), and therefore, again, I am distinct from every physical individual that exists at t^* (because every such individual is either not in *L* at t^* , or is in *L* at t^* but is not a suitable candidate for being me: a brain cell or a small collection of brain cells of UM may be in *L* at t^* but would certainly not be *me*). Taking [1], [2], and [3] in conjunction, what was to be shown is a logical consequence of that conjunction: I am distinct from every physical individual. And therefore also the subject of the present consciousness of UM—the subject of the consciousness of UM at t^* —is distinct from every physical individual (that subject being me). Finally, since I certainly am an individual, it follows that I—the subject of the consciousness of UM at t^* —am a non-physical individual.

In parallel to what was said previously about the normal way of *X* being at a time a soul of *Y*, I note, at this point, that there is also a normal way of *Z* being a subject of the consciousness of *Y* at a given time. This is this normal way:

The normal time-dependent <animal, subject of consciousness>-relationship

- (a') For some time t : *Y* is an animal that is conscious at t .
- (b') For every time t : if *Y* is an animal that is conscious at t , then *Z* is a subject of the consciousness of *Y* at t , and nothing else but *Z* is a subject of the consciousness of *Y* at t .
- (c') There is no Y' distinct from *Y* such that at any time t *Z* is a subject of the consciousness of Y' at t .

If *Z* and *Y* fulfill the condition of normalcy that is specified by the conjunction of the preceding statements, then one can address *Z* simply as *the subject (of the consciousness) of Y*, and of *Y* alone. In fact, this condition of normalcy is fulfilled by the animal UM (for *Y*) and *myself* (for *Z*). Hence it follows that *I am the subject (of the consciousness) of UM*. Am I also *the soul of UM*?

According to the previously stated necessary and sufficient condition for *the normal time-dependent <animal, soul>-relationship*, the truth of the following three statements is necessary and sufficient for *me* being *the soul of UM*:

- (a*) For some time t : UM is an animal that is conscious at t .
- (b*) For every time t : if UM is an animal that is conscious at t , then I am at t a soul of UM, and nothing else but me is at t a soul of UM.
- (c*) There is no Y' distinct from UM such that at any time t I am a soul of Y' at t .

The statements (a*) and (c*) are true (beyond reasonable doubt). All that remains to be shown for establishing me as the soul of UM is (b*). Assume, then, that UM is an animal that is conscious at time t . In order to demonstrate (b*), we must derive from *this assumption* the following five statements (but in doing so, we may make use not only of *the assumption* but also of propositions established on independent grounds):

- (1*) UM is an animal that is conscious at t ,
- (2*) I am a non-physical individual without inherent temporal dimension,
- (3*) I am a subject of the consciousness of UM at t , and
- (4*) if UM is up and about in the time-stretch following t , then I cause some movement of UM after t in the light of the consciousness of UM at t .
- (5*) For every X that is not me: $(1^*)_{X/I}$,² or $(2^*)_{X/I}$, or $(3^*)_{X/I}$, or $(4^*)_{X/I}$ is not true of X .

For from these five statements, it follows, according to D5, that I am at t a soul of UM and that nothing else but me is at t a soul of UM—which is what has to be derived from *the assumption* in order to demonstrate (b*). Now, (1*) is a trivial consequence of *the assumption*. And since the previously stated condition (b') is already known to be true of UM and me (putting UM for Y , and me for Z), we—given *the assumption*—also get (3*) and (5*). This is obvious with regard to (3*), and with regard to (5*), merely consider that *the assumption* and (b') (for UM and me) entail the following: for every X that is not me: $(3^*)_{X/I}$ is not true of X . Moreover, it has already been shown that I am a non-physical individual, hence we already have part of what is stated in (2*).

There remain two questions that are still open:

- (I) Is it true that I am without inherent temporal dimension? (The answer “yes” is required to be correct for establishing (2*).)
- (II) Given any time t at which UM is an animal that is conscious [and, of course, there are in fact many such times t], is it true, then, that I cause some movement of UM after t in the light of the consciousness of UM at t , if UM is up and about in the time-stretch following t [as UM normally is]? (The answer “yes” is required to be correct for establishing (4*).)

In order to give the present section a clear and impressive conclusion, I maintain that the correct answer to both of these questions is “yes” (but my reasons for having this position are presented in the next section and in the section after it). If I am right, then it follows that besides (1*), (3*), and (5*) also (2*) and (4*) have been (rather trivially) derived from *the assumption*,

and hence that the truth of (b*) has been demonstrated, and that, therefore, it is quite true: *I am the soul of UM* (since (a*) and (c*), which are also necessary for this conclusion, have already been established).

3. SOULS AND CAUSATION

Common sense is all in favor of giving the answer “yes” both to question (I) and to question (II) posed in the previous section. Given that UM is an animal that is conscious at time *t*, common sense tells me that, *if* UM is up and about in the time-stretch following *t*, there is no circumstance that could keep *me*—the subject of the consciousness of UM—from causing some movement of UM after *t* in the light of the consciousness of UM at *t*. After all, *without* the almost immediately manifest *effects* of my guidance, UM would almost immediately be *not* up and about after *t*. And common sense also tells me that I do not have an inherent temporal dimension, since I experience myself as not having any temporal parts and as not being a temporal part of anything (not even a point-like temporal part).

But common sense does not seem to be a good advisor with regard to the questions (I) and (II)—because common sense does not seem to present a defensible position with regard to them. How so? The relevant worries can be summed up in two arguments:

Argument 1: If question (I) is answered with “yes”, then it is presumed that I am not an event, nor anything event-like (for example, a temporally specified state of affairs). But then I am not capable under any circumstances of causing anything, since only events or event-like entities are capable of causing something. Now, this negative result makes it (rationally) impossible to answer question (II) with “yes”; it must be answered with “no”. Conversely, if one insists on the answer “yes” to question (II), then question (I) must be answered with “no”. Clearly, one cannot answer *both* questions with “yes” (contrary to what is suggested by common sense).

Argument 2: Previously, the result has been reached—by fairly commonsensical considerations—that I am a non-physical individual. But if this is correct, then I am not capable under any circumstances of causing anything, since only physical entities are capable of causing something. This negative result makes it impossible to answer question (II) with “yes”.

A unified effective response to both these arguments consists in showing the viability—the coherence, or even better: the existence—of *non-physical agent-causation*. In this section and the next I will try to demonstrate the existence of non-physical agent-causation, which attempt, if successful,

takes away the threat Arguments 1 and 2 pose to assuming that questions (I) and (II) can be commonsensically *and* correctly answered with “yes”. As a preface to the considerations that follow, I put on notice that causation—whether of the event-causal or the agent-causal type—is here always taken to be *sufficient causation*, according to which conception of causation a cause, qua cause, is *sufficient for bringing about* its effect (whether or not the cause raises the effect’s probability, and whether or not the cause is a condition without which the effect would not have come about). Note that causal sufficiency need not be an affair of the cause in and by itself: though a cause is here taken to be, qua cause, sufficient for its effect, a cause (in this sense) may be—and commonly is—sufficient for its effect merely in a relative way: on the basis of sufficiency-enabling circumstances (for example, the well-functioning of the mechanical structures involved, this well-functioning not being counted as part of the cause).

Given the general conception of causation employed in this essay, the following is a highly plausible (but *contingent a posteriori*) assumption:

Uniqueness of path of physical event-causation

For any physical event X: if there is a physical event-causal chain that ends with X (on the effect side, not the cause side), then all physical event-causal chains that end with X can be integrated into a *single* physical event-causal chain that ends with X.

One consequence of this is that any two physical events Y and X that are connected by a physical event-causal chain (with Y causing X) are also connected by a *unique integrative* physical event-causal chain: namely, *that* causal chain that begins with Y and ends with X and is a section of *the (unique) complete integrative* physical event-causal chain that ends with X. Another consequence of the preceding principle is that causal over-determination (in the *interesting* sense) is out of the question for physical event-causation (contingently so and *a posteriori*, since the invoked principle, which is the sole ground for this result, is itself contingent and *a posteriori*). In what follows, if there is mention of physical event-causal chains, always *integrative* physical event-causal chains are intended (either complete ones or sections of complete ones).

Now, agent-causation has three possible varieties: it is either *purely direct*, or *purely indirect*, or neither of the two. In *purely direct* agent-causation, an agent causes an event X, and there is no event-causal chain³ which ends with X and the beginning of which is caused by the agent; in *purely indirect* agent-causation, an agent causes an event X, but *only* via an event-causal chain which ends with X and the beginning of which is caused by the agent. It is in principle possible that agent-causation is neither purely direct nor purely indirect; however, it seems to me, *in view of its hardly being realized in reality*, this possibility can be safely neglected. The same distinction that

has just been made with regard to agent-causation can, of course, also be made with regard to event-causation.

How non-physical agent-causation—in both (remaining) varieties: *purely indirect* and *purely direct*—can plausibly enter into the economy of a living conscious animal can be seen in the following way (using UM as a mere example):

Suppose there occurs a certain *voluntary movement* *E* of UM. *E* is a physical event, and there is a (single integrative) physical event-causal chain running backwards in time from *E*. That chain, going backwards, leads to a physical event *E'* in the brain of UM. *E'* causes *E* (purely indirectly).

Now, if *E'* itself were the (counting *in* the direction of time, not inversely to it) *last* link in a physical event-causal chain that does not originate in the brain of UM, then *E* could certainly not be regarded as a *voluntary movement* of UM—contradicting the preceding assumption that *E* is a voluntary movement of UM. Therefore, either (1) *E'* itself has no physical event-cause, or (2) there is a physical event-causal chain that runs backwards in time from *E'* and that *originates* in the brain of UM, which means: the chain has a (counting *in* the direction of time, not inversely to it) *first* link in the brain of UM beyond which it cannot be further prolonged (backwards in time): a physical event *E''*—in the brain of UM—that has no physical event-cause. In both cases, we have a physical event *E** in the brain of UM that has no physical event-cause but that event-causes *E*, the voluntary movement of UM. (It has already been concluded that *E'* causes *E*; and in case (2), *E''* causes *E* because *E''* causes *E'*, which in turn causes *E*.)

Now, if *E** had no cause, then *E** would be a chance-event, and hence the event *E*, being causally based on *E**, could not be regarded as a *voluntary movement* of UM—contradicting the earlier assumption that *E* is a voluntary movement of UM. Therefore, *E** has a cause (although it has no physical event-cause), which cause is, transitively, also a cause of *E*. But who or what is that cause? The interpretation of self-experience, of self-phenomenology knows of only one answer to this question, *notwithstanding* the fact that it has been fed, and also has *digested*, the neurophysiological findings concerning the role of an animal's brain in the causation of the voluntary movements of the animal:

I—that is, the non-physical individual that is the subject of the consciousness of UM (as we already know) *and that does not have a temporal dimension* (there is no indication of it in self-experience)—am a *purely direct* agent-cause of the UM-brain-event *E** (of which fact I am only very indirectly aware); and by causing *E** purely directly, I am, moreover, a *purely indirect* agent-cause of the voluntary UM-movement *E*, in virtue of the physical event-causal chain that begins with *E** and ends with *E*. In fact, I cause *E* in the light of the consciousness of UM at a time shortly before the inception of *E* (say, *E* is the withdrawal of the left hand of UM from this metal pipe, after I hand-felt the pipe to

be getting warmer and warmer rather quickly and anticipated its being burning hot to the hand very soon).

There are alternatives, of course, to this *neurophysiologically informed, self-interpretational* view of the causation of E^* and E . But how plausible are they?

One might, for example, assume that E^* —though not caused by any physical event—is caused by a physical agent (hence not by *me*, since I am a non-physical entity). However, there do not seem to be any physical agents that can plausibly fill that role. But is not the brain of UM a likely candidate for filling it, in fact, the best candidate? It is true that E^* happens in the brain of UM, but E^* is *certainly not* caused by the brain of UM (as little as the movement of an engine's piston, which movement moves the engine, is caused by the engine). E^* is simply something that happens to (in) the brain of UM. And there is also a more fundamental problem with the idea that the brain might be a physical agent-cause of E^* : a *physical* non-event Y can cause an event X only in virtue of some physical event Z that (i) essentially involves Y and (ii) causes X . But we have already found that *no* physical event causes E^* . Hence the brain of UM (though plausibly a physical non-event) cannot cause E^* . By the same token, no part of the brain of UM can cause E^* , nor, for that matter, can E^* be caused by UM itself (which is *very plausibly* a physical non-event).

Alternatively, one might assume that E^* —though not caused by any physical event—is caused by a non-physical event (hence, again, not by *me*, since, having no temporal dimension, I am a non-event). However, there do not seem to be any non-physical events that can plausibly fill that role. This has become apparent only recently, in the wake of the experiments of the neurophysiologist Benjamin Libet. The only likely candidate for being a non-physical event that causes E^* would be an E -directed decision-experience in the consciousness of UM that occurs *before* the inception of E^* : an experience in which I—the subject of the consciousness of UM—experience myself as *effectively* (not just premeditatively⁴) deciding to bring about E . However, the Libet experiments quite unequivocally show that there is no *such* decision-experience: the decision-experience that does occur *follows* E^* , though it is still ahead of E . There is no getting around this, it seems to me, and it is fatal for the view that E^* is caused by a non-physical event. But it presents no problem for the view that E^* is caused by a non-physical agent: *me*; it seems only natural that I become aware of my effectively (not just premeditatively) made decision to bring about E merely after I have effectively made it: by my causing E^* (at the very time of E^{*5}) in order to bring about E (via the neurophysiological chain of command).

All that seems to stand in the way of accepting the existence of non-physical agent-causation is the “*How could . . . ?*” attitude: “*How could*

a cause be non-physical?", "*How could* a cause be not an event?", "*How could* a cause be not even event-like?". The argument implicit in rhetorical questions of this type is, of course, that something—namely, that which is addressed in the questions—does not exist *because* one does not understand *how* it *could* exist. Not a very cogent way of arguing, it seems to me—and especially uncogent if there is positive evidence for the existence of that regarding which one does not understand how it is possible.

Non-physical causation has been attacked on the grounds that it supposedly violates the laws of physics, or, what many erroneously believe to be the same thing, the principle of the causal closure of the physical. The laws of physics, at least, are certainly not violated by non-physical causation; the causal closure of the physical, in turn, is not a law or principle of physics, but a rather shaky metaphysical assumption and certainly far from being a non-negotiable rational obstacle to accepting non-physical causation.⁶

Agent-causation has been attacked on the grounds that it is absolutely incomprehensible, "mysterious" in the disparaging, philosophical sense of the word.⁷ *Much* can be done to alleviate this impression.⁸ And ever since Hume incontestably showed what perhaps should have been obvious to philosophers from the start—that event-causation is *very unlike* the relation of logical consequence—one can rationally choose to find something incomprehensible *in event-causation, too*. Essentially, event-causation is no better off than agent-causation: there is some rational opaqueness to both—which is no good reason to hold that either of the two does not really exist.

The real problem for the existence of non-physical agent-causation is not its alleged incomprehensibility or its alleged conflict with physics. It is the problem that there might be no use for it, that it might be superfluous. True: if something is superfluous, then this does not logically entail that it does not exist. Yet, it must be admitted that the superfluity of non-physical agent-causation, if shown to be a fact, would cast doubt on its existence (we need not invoke Occam's Razor for this).

In this section of my essay, I have argued for the existence of non-physical agent-causation via arguing (1) that I—already shown to be a non-physical individual—have no temporal dimension, hence that I am a *substance* and thus fit for being an *agent*, and (2) that I in fact purely directly agent-cause the UM-brain-event E^* , and purely indirectly, on the basis of E^* (and in the light of the antecedent consciousness of UM), the UM-movement E , *thereby*, by arguing so, rendering crucial support to common sense in the completion of the argument, started in the previous section, that aimed to establish me as *the soul* of UM. My argument in this present section crucially relied on the premise *that E is a voluntary movement* of UM, and hence it crucially relied on the presupposition that there are voluntary movements of animals. Suppose there were no voluntary movements of UM or any other animal. If this supposition were true, my argument, as stated, would collapse.

4. RATIONALITY

Can it be doubted that there are voluntary movements of animals? Can it be doubted that at least some of UM's movements are *voluntary*? It cannot be doubted—at least not by *me* (who is the subject of the consciousness of UM)—that some of UM's movements *feel* voluntary. But perhaps none of them *is* voluntary. How can this be? Easily. It is the case if the following is true:

The coming about of any animal movement is a necessary consequence of the laws of physics and of the physical events that happened before the first animals came into existence.

This—call it “animal determinism”—is just a consequence of the doctrine of universal physical determinism, but it has the advantage that one can believe in it even if one considers *universal* physical determinism to be false (as is strongly suggested by quantum mechanics). If animal determinism is true (and a fortiori if universal physical determinism is true), then there are no voluntary movements of animals (though for some mysterious reason some of these movements *seem voluntary*: to the outside observer, and to the subject of the consciousness of the performing animal). Thus: if animal determinism is true, then my basis for arguing for the existence of non-physical agent-causation is gone. And more than that: if animal determinism is true, then non-physical agent-causation is superfluous, and therefore *non-existent* (“therefore” indeed!—but under the assumption of animal determinism, one is irresistibly drawn to this conclusion, though it does not logically follow). But if there is no non-physical agent-causation, then there are no *souls* (simpliciter), as defined by D6 (taking into consideration what precedes that definition and assuming that any animal is at some time *t* conscious, and up and about in the time-stretch following *t*).

Therefore, if *souls* emerged (and a fortiori if *rational souls* emerged), then the world is more or less bound to be in such a way that animal determinism (and a fortiori universal physical determinism) is not true. Do we have any basis for assuming that animal determinism is not true? Yes, indeed, there is a basis for assuming that it is not true: this basis simply consists in the fact that some animals have *conscious interests*. The fact of conscious interests is not logically incompatible with animal determinism, but it fits ill with it, and animal determinism has nothing to offer to explain it. If animal determinism is true, *conscious interests* are superfluous, and therefore (ought to be) *non-existent*. Under the assumption of animal determinism, one is irresistibly drawn to this conclusion. However, in this case, one cannot rationally accept the conclusion *after all*: the existence of conscious interests is just too plain obvious. This does not bode well for animal determinism.

Conscious interests are embedded in the consciousness that is produced by the brain of any animal. In the indirect sense that can be extracted from the preceding sentence, any conscious animal *has* conscious interests; but the proximate owner of an animal's conscious interests is not the animal, but the subject of the animal's consciousness. Thus, in the case of UM, it is *me* who is the proximate owner of UM's conscious interests. Now, why should this be, if I could bring about nothing in the physical world that would *contribute* to pursuing these interests (typically relating to the physical world)—as must be the case if animal determinism were true?

Conscious interests are a part of any consciousness and make it a locus of (at least) *minimal rationality*: in the light of conscious information about what is the case, however incomplete and distorted, and in the light of conscious interests, however simple, the animal's subject of consciousness is fitting means to ends—at least in a rudimentary fashion (it need not be remarkably deliberate or remarkably intelligent, nor need it be accompanied by self-reflection or conceptualization).⁹ This is the essence of rationality. If having minimal rationality in the sense just described were all that is being required for an animal to be an *animal rationale*—as I would advocate—then human animals would, of course, be far from being the only *animalia rationalia*; rather, being a rational animal would then coincide with being a conscious animal. In whatever way one might ultimately wish to decide this merely conceptual question (i.e., what is to be understood by the word “rational” when applied to animals, including human animals), even *minimal rationality* makes biological sense only if the subject of consciousness of an animal can act on it and make, in the light of it, a difference in the physical world that would not have come about without that subject's action. And minimal rationality *does* make biological sense (it can hardly be denied). Again, this does not bode well for animal determinism.

5. WHY THE EMERGENCE OF RATIONAL SOULS IS A NATURAL EXPECTATION

Suppose animal determinism is false (as I have argued in the previous section), in fact, false on a broad scale (why should there be only a few instances of its falsification?): the coming about of many movements of many animals *is not* a necessary consequence of the laws of physics and of the physical events that happened before the first animals came into existence. In other words, there are many and widely distributed animal movements that *are not* subject to ancient physical predetermination. Consider the complete integrative physical event-causal chains¹⁰ that lead up to these movements. For definiteness, consider a particular animal, AN, many of whose movements are not subject to ancient physical predetermination: they are *the not-anciently-descended AN-movements*, in short: the NAD-AN-movements.

It contributes significantly to the survival chances of AN—by dramatically increasing AN’s adaptiveness—if *most* of the complete integrative physical event-causal chains that lead up to the NAD-AN-movements are short and begin in a central place within AN, where the beginnings of *many* of these chains are subject to survival-directed central control. This is why brains—in particular, AN’s brain—evolved. But what does it mean that many beginnings—in the brain of AN—of complete integrative physical event-causal chains that lead up to NAD-AN-movements are subject to survival-directed central control? First, it means that these beginnings—events in AN’s brain—have a single cause, which is not a physical event-cause (otherwise, they would not be *beginnings* of complete integrative physical event-causal chains); second, it means that that single cause is neither blind nor without wishes nor without a sense of power: it wishes to survive in the situation it is being informed about (normally correctly) and knows (“in its bones”) what it can do in this situation (normally correctly). Its own survival is the survival of AN.

Thus, there is a biological, evolutionary argument for the existence of a single entity that causally controls *many* of the *event-causal origins* (the *event-causal first causes*) in the brain of AN of NAD-AN-movements, *most* of which have, as was noted, an event-causal origin in the brain of AN. This evolutionary argument functions in the same way as all such arguments do: it exhibits how a certain feature of living beings (though not usually a feature whose very existence is controversial) fits naturally into the economy of survival and hence, given the right circumstances, may be expected to appear and be perpetuated in the course of biological evolution. It is strongly suggested by this argument that the causal controller it points to is a non-physical conscious agent. Now, I do not know about AN, but in the case of UM I know this causal controller personally: it is *me*, the rationally active non-physical substance that is the subject of the consciousness of UM, and the soul of UM.

If rational souls exist, then, very plausibly, there has been an emergence of rational souls (in the sense explained in the first section of this essay)—since, doubtless, there is a time at which at least one rational soul has existed, and a time at which no rational soul has (as yet) existed, and therefore, very plausibly, a first time of the exemplification (by something newly existent) of the ontological kind *rational soul*.

This chapter should not end without addressing two objections (*not* of the usual kind). It might be objected that the emergence of rational souls—qua non-physical substances—amounts to a *natural* creation ex nihilo of non-matter *by matter*, and that such a *natural* creation ex nihilo is simply *incredible*.¹¹ Indeed, as things have been described earlier, the coming into the world of souls can be regarded as an *entirely natural* occurrence (that is, one need not invoke for its explanation any special intervention by God): the laws of nature do allow the existence of souls; given the right circumstances, souls come into existence, and stay in

existence because they constitute a survival-asset for animals. In fact, souls simply co-evolved with brains: with the physical organs of animals whose very function it is to serve as the causal vehicles of souls, the non-physical organs of animals.¹² Once one has become accustomed to the idea that *nature* has two sides (a physical and a non-physical one), one does not see why the previously described occurrences of natural history should deserve the designation of *creatio ex nihilo* (of “new stuff” to boot).¹³

It might be objected that the agent-causality of souls cannot be distinguished from the workings of *chance* (and “Why assume agent-causality and souls then?” is the implied skeptical question). This objection is proposed—with a new twist—by Peter van Inwagen.¹⁴ A brief response: If, under indeterminism, the objective probability that UM goes through door A is 0.5, and the objective probability that UM goes through door B is also 0.5 (prior to UM going through either door), then nothing I can do can change that probability; but this does not mean that UM’s going through door A is, or is indistinguishable from, a *chance-event*, in the sense of its having no (sufficient) cause but coming about nonetheless. Of course that event *has* a cause; in fact, it has an agent-cause: *me*; and if I so choose, my *rational* agency will become apparent over, say, 1000 repetitions of my door-choice in a *non-random pattern of frequency*—which pattern, whatever pattern it is (it might be a very artful one), will yet not violate what is required by the objective A-B-probability in the long run.

NOTES

1. Note that *motionless rests* are here also counted among the movements (as limiting cases).
2. The expression $(N^*)_{x/n}$ results from the expression (N^*) by replacing “I” (everywhere in (N^*)) by “X”. If “I” does not occur in (N^*) , then $(N^*)_{x/n} = (N^*)$.
3. The shortest such chain has two links: the event which is the beginning of the chain and the event which is the end of the chain, the first event causing the second. In general, being-a-cause-of is the relation which connects each link of a causal chain to the links that follow it in the chain in the direction of time (i.e., from earlier to later).
4. If I resolve that eight hours from now I shall jump into the pool, then I have made that decision merely *premeditatively*, not *effectively*. If after the eight hours have passed, I say “now” and therewith jump into the pool, then I have made another decision *effectively* (which, however, fulfills, so to speak, the promise of the earlier decision).
5. In agent-causation, *the time of causation* (if, indeed, one insists on requiring such a thing) is always the time of the effect.
6. I examine the causal closure of the physical and defend non-physical causation—of the event-causal and/or the agent-causal sort—in several of my publications; see Meixner (2004, 2006, 2008, 2009).
7. For a typical disparagement of agent-causation along these lines, see Dennett (2003:100).

8. My own efforts in this regard can be found in Meixner (2004: ch. 8 and 9, 2008).
9. For this role of consciousness in the economy of animal survival, compare James (1950: 140–142).
10. For this concept, see Section 3 in this chapter.
11. See for this objection O'Connor (2002: 344).
12. Note that William James regarded *consciousness* itself (in which, according to the present view, souls are embedded) as a causally efficacious organ—a “selecting agency” of the animal, helping it in the struggle for existence; see James (1950:138–139).
13. See also Meixner (2004: 314–323) and Corradini (2008: 205–206).
14. See van Inwagen (2002: 168–175).

REFERENCES

- Corradini, A. (2008). Emergent dualism. In A. Antonietti, A. Corradini, & E. J. Lowe (Eds.), *Psycho-Physical Dualism Today* (pp. 185–209). Lanham, MD: Lexington.
- Dennett, D. C. (2003). *Freedom Evolves*. London: Allen Lane (Penguin).
- James, W. (1950). *The Principles of Psychology*. 2 vols., New York: Dover.
- Meixner, U. (2004). *The Two Sides of Being. A Reassessment of Psycho-Physical Dualism*. Paderborn: Mentis.
- . (2006). Consciousness and freedom. In A. Corradini, S. Galvan, & E. J. Lowe (Eds.), *Analytic Philosophy Without Naturalism* (pp. 183–196). London: Routledge.
- . (2008). New perspectives for a dualistic conception of mental causation. *Journal of Consciousness Studies*, 15, 17–38.
- . (2009). Three indications for the existence of God in causal metaphysics. *International Journal for Philosophy of Religion*, 66, 33–46.
- O'Connor, T. (2002). Libertarian views: dualist and agent-causal theories. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 337–355). New York: Oxford University Press.
- Van Inwagen, P. (2002). Free will remains a mystery. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 158–177). New York: Oxford University Press.