Causal Domains and Emergent Rationality

IVAN M. HAVEL

1. Introduction

There are perpetual attempts in philosophical and scientific literature to bypass the enigma of free will by trying to explain human action on the sole basis of physical causality, perhaps slightly seasoned with pure randomness. Obviously this affects many theorists of rationality and attracts them towards the technical interpretation of the term "rational", as it is used, e.g., in the framework of formal decision theory or, more recently, of artificial intelligence. Such an understanding of rationality has the amusing consequence that most our (human) everyday decisions turn out to be irrational, as various psychological tests have repeatedly demonstrated (Oaksford and Chater, 1998); our decisions in such a framework are rational only in cases when we painstakingly, perhaps mindlessly, execute procedures maximizing some utility function.

In contrast, in his recent book on rationality in action John Searle (2000) presents a conception of subjective, "full-blown rationality" (especially rationality in action) that presupposes, on the side of the decision maker, intentionality, consciousness, temporality, free choice, language, and selfhood. Rationality in this sense is a feature of the decision making process rather than a feature of its result. Therefore "rational" does not necessarily mean "correct", or even "reasonable" in the usual sense: if the decision is made on the basis of beliefs and desires (which then play the role of *reasons*) it is immaterial whether the beliefs are true and the desires desirable. On Searle's account, unlike the traditional philosophical views, beliefs and desires by themselves are not causally sufficient to determine rational actions, rather there is a *gap* between the "causes" of the action in the form of beliefs and desires and the "effect" in the form of the action itself. I shall call rationality in Searle's sense *intrinsic rationality* to distinguish it from *"as-if" rationality* – virtual "rationality" that people often attribute to non-human entities. In this study I present a variant of the concept of rationality, called *emergent rationality*, adopting and extending the notion of emergence as it persists over most of the last century. It is used (with somewhat vague and varying meaning) for objects, properties, or relations occurring at some more observable level of a complex system, and which are supported (some say sustained, caused, or produced) by processes and properties at some other, less visible level, but neither reducible to, nor predictable from, the less visible processes and properties (Nagel, 1961, p. 366).

While the concept of emergence is frequently used in philosophical discussions, the actual functioning of its concrete cases is usually left aside as a task for appropriate sciences. In the present study I shall try to walk the fine line between science and philosophy by proposing a certain tentative way of dealing with the traditional problem of rationality, which is

> How can there be rational decision making in world where everything that occurs happens as a result of brute, blind, natural causal forces?(Searle, 2000, Chapter 1)

Instead of the intrinsic rationality in the sense of Searle, I will depart from the „as-if“ rationality mentioned above. I believe that there are various cases of seemingly rational behavior, both in nature and in the social sphere, where the attribute “rational” cannot be so easily dismissed as nothing but a superficial anthropomorphism. The common feature of such cases is that they arise from complex multilevel systems in a non-reductive manner – which, in a sense, may also hold in the case of intrinsic rationality.

To make this clearer I shall first discuss the view of reality as fragmented into multiple “causal domains” – a partial generalization of a somewhat vague, albeit abundantly used, concept of a *level* (e.g., level of description). This makes it possible to conceive of emergent rationality as a phenomenon based on the interaction of (at least) two different causal domains, in one of which, for example, a nontrivial selection process is realized that, in the other domain, yields some sort of effectively “rational” behavior.

Such a domain-oriented approach may provide a framework for posing some questions of particular interest, such as whether, and which type of, rationality can be ascribed to collective systems of units commonly considered as *non-rational* (i.e. systems like neural networks or robot societies), and whether, and which type of, rationality can be ascribed to collective systems of *rational* individuals (systems like human societies or internet communities).

The reader who expects a solution to the problem of rationality, or at least a treatment based on precisely formulated concepts and arguments, may be somewhat disappointed. Indeed, my intention is no more than just to provoke some new thoughts and offer themes for discussion. Moreover, I could not avoid a certain sort of ambiguity in my treatment so that the reader is free to alternate between three types of reading: the psychological, the epistemological, and the ontological.

2. Four Test Examples

For the sake of easier discussions I shall first present, as a preview, four illustrative examples of particular capabilities of complex systems – complex in the sense that they inherently involve multilevel interactions. I shall later discuss the extent in which such capabilities may be viewed as cases of emergent rationality.

First example: The Clever Fluke. There is an often quoted case of the fluke *Dicrocoelium dendriticum* as an example of a parasite manipulating an intermediate host to increase its chances of ending up in its definitive host. The definitive host is a sheep, and the intermediate host is an ant. The normal life cycle of the fluke calls for the ant to be eaten by the sheep. To achieve this the fluke changes the ant’s behavior by “cleverly” manipulating its neural system.[1] Whereas an uninfected ant would normally retreat into its

1. In fact, the reality is more intricate: the neural system is entered by one or two self-sacrificing fluke specimens from a group of about 50.

nest when it became cold, infected ants climb to the top of grass stems and remain immobile. Here they are vulnerable to being eaten by the fluke's definitive host (Dawkins, 1982, p. 218).

Note that the fluke species can be viewed as a quasi-material, diachronically evolving entity that is materialized in scores of individual fluke specimens. The ant-manipulating behavior is, in a sense, a property of the fluke as a species, a property with a certain distribution in a population of flukes, as well as an actual behavior of any particular fluke specimen). In the common language these distinctions are not always explicit; they become relevant for the analysis of certain emergent phenomena, as will be seen later on.

Second example: The Expressive Language. Our second example of a "clever" multilevel system is any natural language. Consider, for instance, the English language with its ability to express a variety of sophisticated ideas, such as the expression of acts not as facts, but as contingencies, as in the sentence: "There are no odds at which I would bet my life against a quarter, and if there were, I would not bet my child's life against a quarter."

We might want to differentiate between, first, particular successful speech episodes, second, the linguistic skills of the speaker, and third, the "intelligence" of the language as such. Of course, without speakers and their speech acts language would not evolve (let alone exist), and without language the speakers would be silent.

Third example: The Chess Machine. Computers have played grandmaster chess since 1980's and IBM's "Deep Blue" machine defeated Garry Kasparov, probably the best human player ever, in May 1997 match. Several times during the match, Kasparov reported signs of mind in the machine (Moravec, 1999).

This example may invoke the long-standing dispute about mentality of machines. In our context the most important distinction is between the performance of the chess playing program and the intellectual activity of people who developed the program.

Fourth example: The Rational Mind. Assume that you received a letter from organizers of a conference asking you to give a lecture on a topic of your choice. Three days later you answer affirmatively and give a title. What happened during these three days? You were probably considering various alternatives and in the course of arriving at the decision you were, I presume, quite certain, that your decision was arrived at freely, consciously, and entirely on the basis of mental reasoning. You could feel certain pressures but you would be fully aware the whole time that a different decision could have been arrived at, perhaps even one that aborted the decision making process entirely. So, I believe, this is an apt example of rational decision-making in Searle's sense.

But this is not the end of the story. You thought about it and realized that there are no sound arguments against the claim that all your

> mental phenomena are caused by neurophysiological processes in [your] brain and are themselves features of the brain,

as Searle puts it (1992, p. 1). In fact, there exists a vast amount of scientific data about processes in the brain at the neural level (and perhaps some lower levels), but there are only speculations about how these processes might support, sustain, control, cause, or create an illusion of, mental activity.

I am now going to explore the above examples in some detail – not as to their individual nature but each as a representative of a large category of more or less similar systems. Let us first observe three general features they share and one in which they differ:

1. Each of them is a complex dynamical system that we, as observers, are used to describing and comprehending at two or more substantially distinct "levels" (phylogeny vs. ontogeny, diachrony vs. usage, programming vs. performance, mind vs. brain). The study of events at each particular level typically requires a distinctive scientific approach. I will discuss and generalize the concept of a level in Section 4.
2. The events at one level are in one way or another importantly interlinked with, or dependent on, events at another level (or at several other levels). The nature of this mutual *interlevel interaction* is not always well understood and may even require a revision of the traditional notion of causality. I will say more about this in Section 6.
3. Each of the cases exhibit a certain type of *purposive*, *intentional or rational behavior*, at least if we allow the "as-if" interpretations of these words. Such behavior is commonly attributed to entities (organisms, agents, persons, etc.) with respect to a single level, but as our examples suggest, it may be causally efficacious on a different level than the level that produced or sustained it. I will return to this point in Section 10.

In one interesting aspect our examples differ. Except the first example the systems involve, at a certain level, conscious beings able to act, in principle, of their own free will.[2] The difference is, however, in the actual role of the conscious and free beings and in the level on which they act. In the case of the Expressive Language the relevant activity of individual speakers occurs at the lower, "finer" level of the system while in the case of the Rational Mind consciousness is associated with the higher, "coarser" level of the system. The case of the Chess Machine differs even more: the programming activity is intentionally directed to the machine performance and hence it is not quite appropriate to talk of "levels" in this case (the same holds for other designer-artifact systems).

3. About the "As-If" Nature of Some Statements

Often people use various anthropomorphic terms, like "rationality", "purpose", "desire", "intentionality", etc., about behavior of animals, machines, social institutions and other entities that behave in an appropriately sophisticated way. In most cases such usage expresses just a tacit feeling that if we people were in the position of such entities, we would (consciously) behave in a similar way. Most philosophers cautiously indicate the metaphorical sense of the words by adding the particle "as-if" (or equivalent).

The general attitude is that the metaphorical uses of mentalistic terms are nothing but linguistic conveniences that express the beholder's external view. I propose to consider,

2. This statement is not quite substantiated because I did not made any metaphysical assumption about which entities are really conscious and free and which are not. If the reader counts flukes to be conscious, or natural-language speakers to be zombies, I have nothing against it. We would only have to look for other examples.

instead of the fact *that* a certain entity appears rational, purposeful, intentional, etc., rather *what is behind* such appearance – is there something intrinsic to the nature of the entity in question that makes it look like having such properties?

Most authors recognize only two alternatives, either the intrinsic ascription, or the "as-if" ascription, and often they are quite decided about what is what. There are few who propose further options (e.g., Haugeland, 1998). Expectedly, the main stimulus for studying the intermediate cases comes from biology, which offers unlimited number of impressive cases of apparent rational behavior.

Let us recall the statement "To achieve this the fluke specimen changes the ant's behavior …" used in the description of our first example. Except for scrupulous scientists most people would accept the phrase "to achieve this" in this context as a legitimate and innocently metaphorical phrase used just to make the reading easier (the same holds for another, even more frequent phrase: "in order that"). We usually do not presume that such teleologically flavored expressions, when referring to (lower) animals, could be understood in their literal sense since we generally do not expect inherently intentional and conscious behavior from such a diminutive creature like the fluke.

But is this all that can be said about the phrase "to achieve this" when it is used in non-human contexts? Perhaps we might try to give it a more informed interpretation, namely, that it could possibly reveal, instead of somebody's (in our case the fluke specimen's) intention, an influence from, or at least existence of, some other, currently "invisible" level. In the particular case of the Clever Fluke it could well be the level of biological evolution. Note, however, that the teleological discourse, removed from the level of specimen behavior, may inconspicuously move to the evolutionary level where it might obtain the form of phrases like "natural selection favored this or that strategy". Isn't it so that the words "selection", "favor", "strategy" have a certain vestige of teleological meaning? Well, even here perhaps the teleological aspect can be removed by referring to the blindness of the entirely causal Darwinian variation-selection-replication process.

I was carelessly playing here with various options – in fact, with three of them: intentionality of individual animals, intentionality of evolution, and no intentionality at all – for a purpose. I wanted to demonstrate, first, that levels of description (soon to be generalized to causal domains) are relevant to the problem of rationality; second, that the attribute "as-if" should be relativized to particular levels; and third, that the nature of the sources of certain phenomena can be left open if our scope of interest is restricted.

In the following I will differentiate between *"as-if" rationality* (indicating the conviction that "nothing is behind" the phenomenon) and *apparent rationality.* I will use the latter term without, or prior to, any judgement about whether we are dealing with the case of intrinsic rationality, "as-if" rationality, or emergent rationality (to be introduced later). For example, I can ascribe apparent rationality to a concrete individual act of another person (perhaps even to myself) if the same type of act can be performed consciously (deliberately) as well as unconsciously (compulsively).

4. Causal Domains

In the preceding sections I frequently and in important contexts used the term 'level'.

Now I will generalize it in a suitable way for our purposes. In scientific as well as philosophical literature there are frequent references to concrete instances of levels, often piled up in one or another hierarchy or at least treated as if one level were "above" the other. Thus, for instance, we often read about the atomic level, molecular level, cellular level, neural level, etc., up to the psychological level, behavioral level, and societal level. However, for the most part we are left alone with our intuition about the very *concept* of a level. Judging from the usage, the only general, common feature of all levels is their epistemological meaning: they indicate much better understanding, on the side of experts, of relationships and laws *within* a particular level rather than *between* different levels.

I prefer to use the term *domain* instead of the term *level* to suppress the tacit assumption of the existence of some underlying hierarchy of levels (where some levels are usually called "upper" and some "lower"). Thus the concept of a domain is substantially more general than the concept of a level. In fact, even our everyday experiences present the world as if it were broken into various domains, each suitable for a certain type of interest and a certain type of action, and each somewhat better understood if taken separately than if combined with other domains.

In the scientific enterprise as well as in everyday life, *understanding* of the interrelationships of spatio-temporal events usually means the ability to *explain* and *predict* something. Explanation and prediction is based, in science, on *causal laws* (known or unknown), and in everyday life on the common-sense idea of one event (or state of affairs) – a *cause* – bringing about another event (or state of affairs) – an *effect* – not just by coincidence. In view of this, I shall first deal with domains based on causal relations (until Section 8 I often omit the adjective "causal").

According to the traditional view, causes are events antecedent to their effects. Moreover, some theoreticians maintain that whenever one event causes another, it does so in accordance with a general law.[3] The following introduction of the concept of a causal domain (the term was used casually by Kim, 2000, p. 69) deliberately presumes causation in the narrower sense (sometimes called the left-to-right causation). More general uses of the term "causation", e.g., the "bottom-up" or "micro-macro" causation, will be discussed in Section 6 as a type of extrinsic relation between domains.

Let us call a *causal domain* any segment (or fragment, or component) of reality within the scope of which causal relations appear to be (i.e., are presented to our knowledge as) *manifest* (obvious, apparent), *comprehensible* (intelligible), and *mutually coherent*. Or, more appropriately, they appear to be *more* manifest, comprehensible, and mutually coherent than causal relations *between different* domains are. This formulation is not meant as a rigorous definition – I am just trying to characterize the existing intuition, so the circularity should not be a hindrance. The idea of causal domains is admittedly vague[4] (depending partly on the vagueness of the underlying concept of causality), so let me clarify a little what I mean (cf. Figure 1.).

1. I do not presume, in general, that observing *concrete* instances of *causal relations* (let us call them *causal episodes*) automatically implies knowing the corresponding

---

3. In this paper, I am not concerned with the issues of the nomological character of causality.
4. The vagueness of the idea may not impede the fertility of some of its applications.

causal laws,

2. by saying that these relations *manifest themselves* I simply mean that in concrete situations they are easily recognizable and identifiable as causal,
3. by saying that they are *comprehensible* I want to imply that most of us accept them without a need for further explanations, or without surprise,
4. *mutual coherence* of the rel
ations means that they belong to one common, apparently consistent, *causal network.* In particular, various causal episodes may fit together and form arbitrarily long *causal chains.*

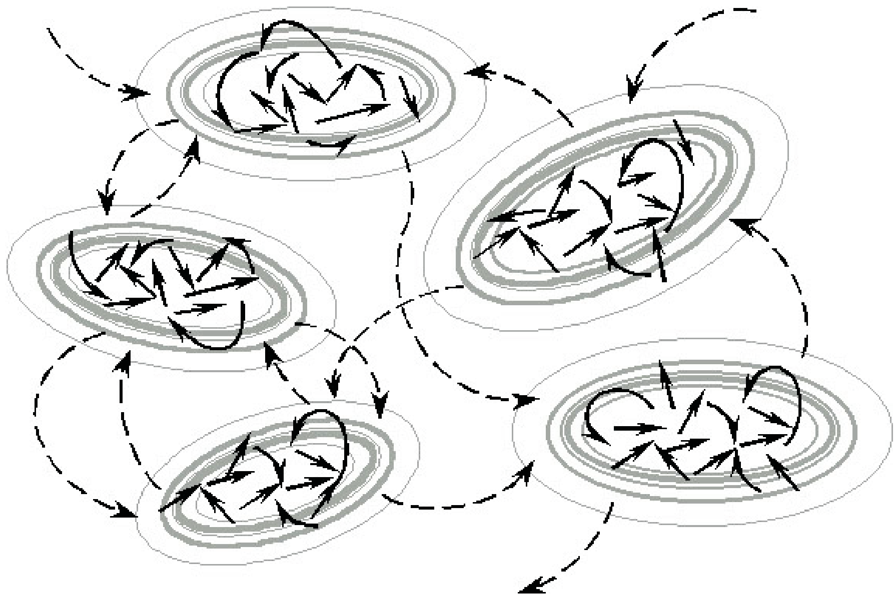

Figure 1. Causal domains. Causal relations within a domain are more manifest, comprehensible and mutually coherent than between different domains.

The concept of a causal domain can be used with respect to any individual (private) knowing of the world, as well as to the collective, scientific knowledge. Examples will make it clearer.

On the individual side, we have our everyday experience of temporarily restricting our attention to the (presumed) causal context of an activity that we are engaged in or of a problem we want to solve. Let us say, for example, that you have a problem with the engine of your car. In order to fix it you must take into account the various causes and effects relevant to the functioning of the engine. But you would probably not take into account, say, the causes and effects relevant to the chances of it raining the next day. Later, though, when you cut your lawn, it may be the other way around. Note that such "causal domains" are more or less shaped by your individual knowledge.

More interesting examples of causal domains are the particular subject areas of natural science (like, for instance, quantum physics, molecular biology, or evolutionary biology) and the fields delimited by different research methods. More or less anything we find in scientific and philosophical discourse designated or recognized as a "level" falls under the concept of causal domain. In scientific contexts causal domains and levels of description are typically shaped by collective knowledge.

Each causal domain may be viewed as a world unto itself: it has specific individuals, universals, properties, aspects, relations, laws, etc. Such things may be peculiar to one particular domain – then I shall call them *endemic* to the domain. Other things may belong to, or be meaningful in, several, perhaps many, domains (so-called *multidomain* entities and *shared* concepts).

A few examples of the latter case may be useful. First, consider a natural physical phenomenon like lightning. Whether as a token (concrete event) or as a type (the generic concept), it can be conceived as a multidomain entity. It, so to say, "penetrates" through many domains differing in scale: from the macroscopic scale, as we see it in the sky, to the microscopic scale of its physical description as a collective flow of charged particles. The difference in scales is more than twelve orders of magnitude which makes it difficult (for an observer) to comprehend this phenomenon as one single entity. Another, more interesting example is the mental state of, say, fright. It may be studied in the behavioral domain (as a pattern of behavior of most animals), in the endocrine domain (as a release of adrenalin), in the mental domain (not quite pleasant first-person experience), in the genetic domain (as a trait carried on certain genes), in the evolutionary domain (as a factor in natural selection), etc.

Such multidomain entities may be contrasted with cases of generic concepts and properties *shared* by various domains, i.e. having analogous, or even the "same", meaning in various domains, partly due to their generality, partly due to the limitations of our language. Obvious examples are the concepts of space, time, causality, and most mathematical abstractions. Many domains, especially those that are a by-product of science, are not directly accessible to our intuition (think, e.g., of the domain of elementary particles). We can describe them only mathematically or by metaphorical transfer of vocabulary, borrowed from perceptionally accessible domains (e.g., the "spin" of elementary particles or the "collapse" of a quantum wave function).

A typical causal domain lacks sharp borders, and what we count or do not count to it depends on how far we are able to extend the connected network of mutually coherent causal relations. On the individual side it is often related to how far our "sight" is able to reach. For a particular scientific discipline this "sight" is not so much dependent on the visual field and viewpoint of a concrete observer but rather on concepts, quantities, laws, and paradigms that are peculiar or significant for the current state of knowledge in this discipline. Naturally, such a discipline is limited, but at the same time it lacks a clear border. So, instead of a border, we deal with the concept of a *domain horizon*.

This brings us to the challenging question: are causal domains really "out there" in the world or not? I think both, in a sense. There is something inherent in the fabric of reality that makes it easier for us to cope with the world in relatively separated regions (levels of description, areas of interest, spheres of knowledge). At the same time there is something inherent in human nature (limited scope of conscious attention, physiological limitations of perception, etc.) that forces us to approach the world in fragmented way. Thus, to the same extent that we are realists about the difference between trees and forests, water drops and clouds, bees and beehives, neurons and brains, and causes and their effects, we should be realists about causal domains. Yet, undoubtedly, we have a great amount of freedom to fix the details of such decomposition. Our picture of the world is a dynamical outcome of a never-ending circular hermeneutic process: our world is *enacted* (Varela *et al.*, 1991).

Let us imagine that a certain collection of causal domains can be ordered into a linear sequence with the help of some natural or artificial ordering characteristic. For example, we may sort the domains according to the dominant spatial and/or temporal scale of typical objects or processes in each domain (Havel, 1996), according to the structural (mereological) subordination or functional dependence of entities belonging to different domains (Scott, 1995), or according to any other suitable characteristic of the domains in question. Only when the collection of domains is so ordered, is it appropriate to talk about a *hierarchy of levels*.[5] The term *level*, used for any of the domains in the hierarchy, is therefore relative to the chosen ordering characteristic. In our account the term "level *of description*" is derivative and the difference between *higher* levels and *lower* levels is then implied. So, for example, in biology we can talk about a large hierarchy of domains ranging from the molecular level, through the level of individual organisms, up to the level of the biosphere. Or, in the time-scale hierarchy, we can separate the phylogenetic domain, the ontogenetic domain, and the domain of behavioral episodes. On the other hand, the manner of treatment of the mind as a "higher level" and the brain as a "lower level" in a certain hierarchy is, from the point of view explained above, somewhat dubious: there is no obvious ordering characteristic applicable there (in fact, this is one of the reasons for my preference of the concept of a causal domain).

We are usually able to shift our focus from one domain to another: in daily life we do it all the time; in science it may depend on profession: scientists in one field just don't feel quite "at home" in other fields. At the same time it is difficult, if not impossible, to keep several domains in sight simultaneously when they are sufficiently "distant" from each other, while "neighboring" causal domains cannot be sharply separated.

5. The Mental Domain

We are used to ascribing causal nature not only to physical relations (in the broad sense of the word "physical") but also to relations that involving our mental states. It is therefore appropriate also to take into account *mental causal domains.* For example, suppose you become frightened after seeing a snake and that makes you freeze up. You might sense this episode as a causal chain and describe it in a causal language, as if your perception directly caused your fright and your fright directly caused your immobility. No allusion to the physiology of your body is required.

The subjective mental domain (or "inner world") of a person includes experienced mental states and their causal relations (when something causes a mental state, or when a

5. To simplify matters we do not consider here abstract "hierarchies" formed by mutual inclusion of domains.

mental state causes something else). Hence, it also includes the mental representation of that part of the physical world that can be affected by intentional acts, and also affect mental states, of a person. In particular, it includes the person's body and bodily movements.

We can distinguish (with, e.g., Chalmers, 1996) two kinds of mental domains, the *phenomenal domain* (of first-person experience accessible to consciousness) and the *psychological domain* (the causal or explanatory basis for behavior described in the third-person way). Within the phenomenal domain we can further differentiate the perpetually changing subdomain of consciously attended mental states from the realm of the unconscious. The assumption that conscious beings have phenomenal mental domains that differ from each other only in their concrete contents (the experienced instances of mental states and experienced instances of causal relations) allow us to talk "objectively", or more appropriately, "intersubjectively", about the mental domains of other persons.

An instance of a causal relation in the phenomenal causal domain is undeniably "true" *within* that domain in a similar sense as a hallucinated object is undeniably "seen" by the hallucinating subject. Consequently, it may not be compatible with causal relations in the psychological or other domains.

6. Connections Between Causal Domains

So far I have been somewhat reticent about the ways events in one domain could influence events in another domain – except for the vague consequence of the "definition" (of causal domains), according to which if the influence between domains is causal then it is, in general, less manifest, less comprehensible and less coherent than causal influences *within* each domain. This does not mean, however, that the interdomain relationships should be considered weak or irrelevant. In fact, one of the motivations behind our approach is to allow for certain kinds of efficacy between domains that may perhaps have a different nature than the classical causal relations (and laws).

Let us see what present-day science has to offer. In the past few decades, scientists have dealt with various situations that characteristically involve two or more different levels (or domains). Besides already mentioned statistical physics, there are theories of structural and/or shape interaction (e.g., of large molecules), quantum effects (e.g., non-local quantum phenomena), and various cooperative, non-linear, chaotic, synergetic and emergent phenomena. For some types of influence we lack (at the current state of knowledge) a formal description or even an intuitive grasp. The most peculiar case is, of course, the psychophysical (mind-body) interaction.

First, let us make some terminological distinctions. I will use, in general, the term *efficacy* for any conceivable influence or dependence, without any apriori claim about its physical (or nonphysical) nature and even about its direction. Thus a connection between different domains is efficacious if, in virtue of it, events in any one domain can bring about events, or affect the actual state of affairs, in the other domain. Hence, (interdomain) *causality* is a special case of (interdomain) efficacy. In the case of causality we automatically assume an explicit direction of effect; whenever I also want to include reciprocal or mutual efficacy (whether causal or not) I prefer to use the more general term *interaction*. Kim (1974/1993) analyses various cases of what he calls "noncausal connections" as being dependent on the structure of events or states of affairs in question.

Among *non-efficacious* connections between domains we may list some *logical* or *analytical relations* that often depend on the way we understand what the nature of "events", "state of affairs", "properties", etc. The important concept of *supervenience* is defined by some authors essentially as a logical relation (Kim, 1974/1993), while some other authors treat it as just a feature of causality (Searle, 1992, pp. 124-126). The term *correlation* and *parallelism* may be considered to be tentative words for phenomena currently lacking a causal explanation.

Across this classification are *systemic* connections based on the existence of a complex multidomain system (cf. Section). For instance, the important phenomenon of *emergence* can sometimes be treated as a directed causal relation and sometimes as a non-efficacious and timeless systemic relation (the type of the treatment depends on our point of view).

Due to the nonexistence of a precise concept of an extension of a domain there is no exact dividing line between *endemic* causality and *interdomain* causality. The endemic causal episodes often have temporal character (the cause precedes its effect; hence the "left-to-right" intuition for this kind of causality). Interdomain causation typically involves at least two different domains that are often viewed as different levels in a certain scalar hierarchy. The typical cases of interdomain or interlevel causality are *upward* causation and *downward* causation.

When we deal with an efficacious relation between different domains (or, more commonly said, between different levels), we have much to learn about its nature before claiming that it is causal, or even that it is an instantiation of some general causal law. We should not take it for granted that ideas used in thermodynamics for explaining heat or in evolutionary theory for explaining mimicry can be applied everywhere, including the theory of mind.

7. Biological Naturalism and Causal Gaps

John Searle (1983) formulated his *biological naturalism* as the thesis that "mental states are both *caused by* the operations of the brain and *realized in* the structure of the brain." This leads to the interpretation of causal sequences at different levels as "not independent causal sequences, but the same causal sequence described at different levels" (Searle, 2000, Chapter 9).

In the framework of causal domains we cannot say so easily "*the same* causal sequence", at least in cases without a clear theory of interdomain connections. Perhaps we could be more cautious and render the situations that Searle had on mind as cases of "causal parallelism": certain causal episodes in one domain X appear to be accompanied by certain causal episodes in another domain Y.

Consider, for instance, that your percept of a snake caused your fright (causal relation in the mental domain). In parallel, as if by coincidence, a certain pattern of neuron firing in your visual cortex caused some other pattern of neuron firing in your motor cortex, inhibiting your muscular movement (an instance of a causal relation in the neural domain). We don't know yet what the *nature* of the seemingly efficacious interdomain connection is, nor even its direction (or ever if it has a direction at all), yet the *existence* of a connection is a sound scientific hypothesis. Talking about causal parallelism in such a case is scientifically vague but intellectually helpful.

In fact, it leads to a difficult issue of making sense of the possibility of *free* decision making in the mental domain. Following up our snake example, consider that when seeing the snake you overcome the instinct and launch a conscious deliberative process aimed at a decision about your future behavior, for instance whether to stop moving or run away (there may be good *reasons* for both options). Is the idea of causal parallelism still applicable?

John Searle (2000) elaborated a theory of intrinsic rationality in action presupposing, on the side of the decision-maker, conscious awareness of the existence of alternatives for a free choice. Thus, instead of being causally determined by an antecedent set of beliefs and desires, rational decision making presupposes a *gap*. In Searle's words, it presupposes

> […] a gap between the set of intentional states on the basis of which I make the decision, and the actual making of the decision. That is, unless I presuppose that there is a gap, I cannot get started with the process of rational decision making. […] We presuppose that there is a gap between the "causes" of the action in the form of beliefs and desires and the "effect" in the form of the action. This gap has a traditional name. It is called "the freedom of the will" (Searle, 2000, Chapter 1).

Searle's account of the human rational process is primarily presented within the framework of the subjective mental domain of a person. In that domain, the gap can be viewed as an opening, or play,[6] for some sort of external intervention––the free will (or what is subjectively sensed as the freedom of the will). The actual "sources" of free will are beyond the horizon of the mental domain, and thus the options are open whether to take it as a primitive principle or to believe in some natural explanation, perhaps exposed to scientific investigation. In Figure 2 there is a schematic "parallelogram" illustrating the situation (cf. Searle, 2000, Chapter 9.)

---

6. Let me note that the Czech word 'vùle' means both 'will' and 'play' (= clearance).

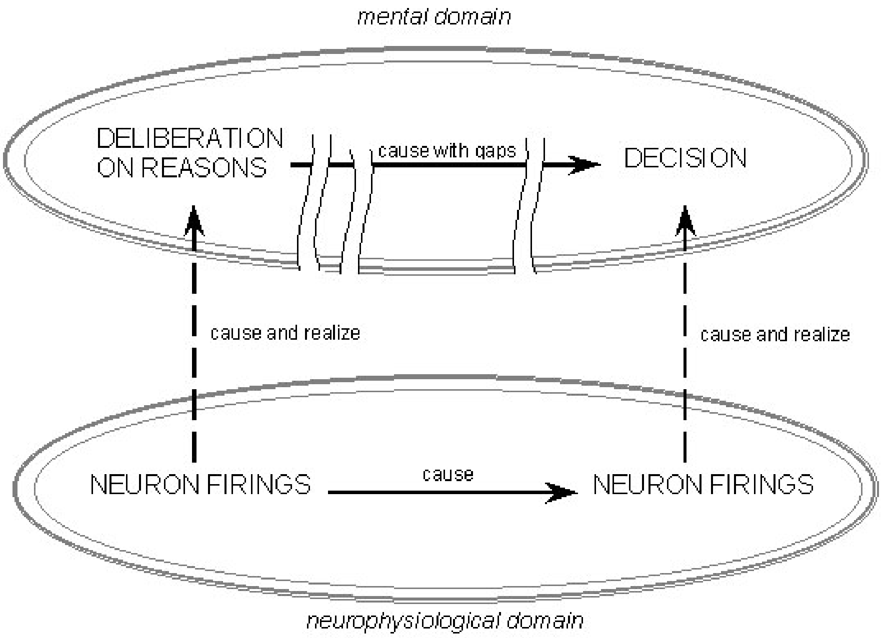

Figure 2. Searle's parallelogram with experienced gaps. (According to the compatibilist hypothesis gaps appear only in the mental domain, neurophysiological domain is deterministic.)

In the scheme presented in our framework the phrase "external intervention" may actually mean "intervention from another causal domain." This, incidentally, may help to sort out various physicalistic and reductionistic theories. What fills the gap? Searle's answer, "Nothing", can be interpreted as: "Nothing in the mental causal domain." This opens the question of the nature of interactions with other domains, which either may or may not fill the gaps deterministically.[7]

I am not going to entertain this issue here, and I mention it just to motivate the idea of extending the concept of a gap to arbitrary causal domains. Let us define, rather vaguely, a *gap in a causal domain* as any opening in the network of causal relations (in that domain) for efficacious influences from another domain(s). Schematic picture of a gap is in Figure 3.

7. In the last chapter of his book (2000), Searle proposes two possibilities of explanation of human rational behavior (psychological indeterminism either coexists with, or is matched by, neurobio- logical indeterminism).

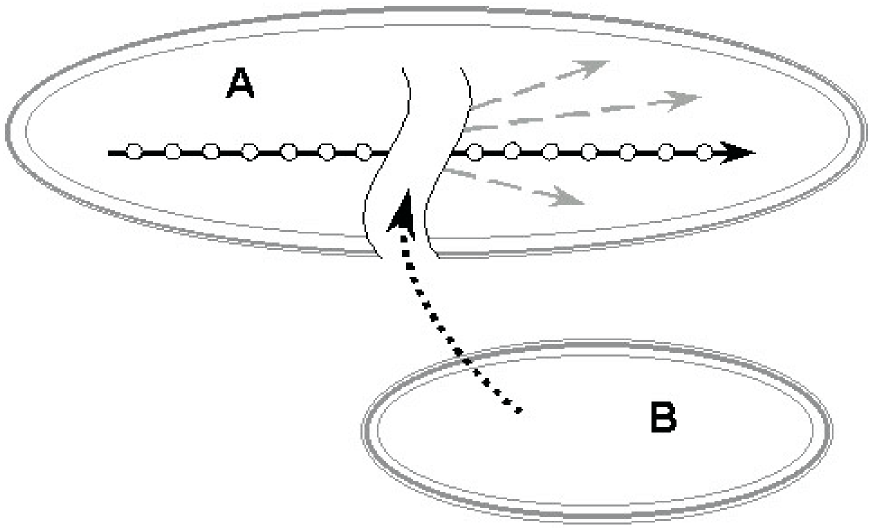

Figure 3. Causal domain with a gap: an opening in the network of causal relations in domain A for influences from domain B.

For example, in the case of the fluke we may ask what is the actual origin of the fluke's solution to the nontrivial problem of reaching the sheep's digestive organs. This question may lead us to the evolutionary domain, but there we can find a "gap" in the linear flow of causally connected events: a "play" where Nature (or Evolution) could make a "choice" among a practically infinite number of alternative solutions to the fluke's problem. The gap has been bridged by blind chance (this is one point of view) or by rational design (another point of view).

From one's "view" within a domain, gaps appear, so to say, on the horizon of the domain (cf. Section 4). Under a suitable extension of the domain, it is possible that some of the gaps would cease to exist while others might emerge. Thus the concept of a gap is always domain-relative.

For explanatory purposes within a particular domain, such gaps are usually "filled" with the help of various default assumptions. In science it is often the assumption of randomness. For example, biologists, when they discuss neo-Darwinian theory in the evolutionary discourse, would make the assumption of random occurrences of mutations random (read: the gaps in evolutionary domain are filled with random events). This sidesteps looking for another causal domain in which each particular mutation could be explained as the outcome of a certain causal chain. Similarly we may read Searle's theory of gaps as a way of saying that the default assumption for the gap filler in the mental domain requires the concept of the self.

8. Causal Domains, Explanatory Domains, and Rational Reasoning

Let us refer to the following observation (by Kim, 1974/1993):

> It is congenial to the broadly realist view of the world that most of us accept to think of the network of causal relations in the world as underlying, and supporting, the network of explanatory and other epistemic relations represented in our knowledge of it.

In particular, we expect many "why" questions about events in the world to be answered by pointing, at least indirectly, to their causes (rather than to their effects). So, for example, when asking: "Why do pebbles have a smooth surface?" we will be satisfied with the reply, "Because frequent collisions with other pebbles stripped them of projections on the surface." We might be less satisfied with the reply: "Because a smooth surface helps them to persist in the stream," or with the trifle: "Because they are defined so" (even though these answers are true as well).

Now let us recall the main aspects of causal domains, namely, that causal relations *within* domains are manifest, comprehensible, and mutually coherent more than causal relations *between* domains are. If our knowledge of the world is structured into such domains, then the "network of explanatory and other epistemic relations" is also structured into such domains. Thus we can introduce the concept of *explanatory domains* within which the explanation of concrete events and facts is easier, more direct, or more acceptable than explanations between explanatory domains. In the first approximation, if we consider only the ordinary, left-to-right type of causality and the corresponding (let us say "right-to-left") explanations, there is a relative match between causal domains and explanatory domains. This match may be understood as a result of a continuing mutual interaction of our ontological views with our epistemic conceptions.

There are, of course, other kinds of explanatory relations. Some of them may refer to various types of interdomain connections; their explanatory force (subjectively valued) obviously depends on our acceptance or nonacceptance of the type of interdomain connection referred to. Other, particularly interesting explanatory relations may refer to voluntary (human) behavior and accept explanations that cite *reasons* rather than just causes.

Evidently, the explanatory success of an answer to a "why" question concerning a voluntary action may crucially depend on the characteristics of the inquirer's explanatory domains, and the veridicality of the answer may depend on the agent's choice of a particular domain as a background for reasoning about the action. Indeed, talking about reasons is meaningful only if they are somebody's reasons, and that the "somebody" has to be a free entity with intentions (a *self* in the Searle's sense or an *existence* in the Heidegger's sense). Only then, the argument would go, can we expect the ability to weigh various pros and cons of alternative choices from such an entity.

To develop a notion of explanatory domains that would include domains of rationality (i.e. causal domains in which, besides causes, reasons are also manifest, comprehensible, and mutually coherent) appears to be a nontrivial assignment. Fortunately, our task is somewhat easier since the types of *apparent rationality* we are concerned with allow us to assume the intersubjective position.

9. Multidomain Entities and Systems

I have already mentioned that there are objects and events that can be treated as multidomain entities. Our previous examples (of lightning and fright, cf. Section 4) only demonstrated the relevance of various domains for the description of such entities. There are, however, cases for which various domains play an essential role from the functional point of view. The term *multidomain system* (and its particular case, the *hierarchical system*) might therefore be more appropriate, whether it is used in the ontological sense or in the epistemological sense. In fact, it is often the case that the collection of epistemologically relevant domains of such a system (e.g. the hierarchy of levels of description) more or less coincides with the collection of ontologically relevant domains (the hierarchy of organizational and functional levels).

Particularly interesting for us are multidomain systems that are complex, so to say, twice over: first, they are complex with respect to each particular domain relevant for the system, and second, they are complex due to the presence of a web of multifarious efficacious interactions between these domains. Not just the presence, but the durability of the whole system may crucially depend on these interactions. The systems and entities to be discussed here typically evolve over time in various ways (depending on which domain is considered). For simplicity, however, I will not pay particular attention to their origins.

Obvious examples of complex multidomain systems are living organisms, ecosystems, social organizations, or computer systems. In this section I reconsider the examples from Section and present each as a multidomain system emphasizing the types of interdomain interactions that are for one reason or another interesting.

First example: The Clever Fluke (revisited). In the case of the fluke, two particular domains of biological discourse are clearly involved: the *evolutionary domain*, in which the main objects of study are animal species (and their evolution), with underlying time scales of millions of years and more, and the *domain of individual animals*, with underlying time scales from seconds to years. For some purposes another domain, the *domain of populations*, can be distinguished (with mostly statistical characteristics). The domain of individual animals can be further divided to the *ontogenetic* (or *developmental) domain* of individual life (months and years) and the *episodic behavioral domain* (seconds and days). One could (I will not do it) further add various physiological and biomolecular domains supporting the behavior.

The “distance” between the evolutionary domain and the domain of individual animals is so huge (especially due to different time scales) that changing the discourse from one to the other requires a radical mental shift. When we pay attention to one of the domains, the other almost disappears from our sight; consequently, we are not confused when the same words are used in both domains. So, for example, the statement “On islands smaller animals grow, while larger animals shrink” is logically meaningful in both domains (while its factual meaning is rather different).

Let us consider, theoretically, a conglomerate of domains relevant to the fluke strategic behavior, from the evolutionary domain of the fluke species to the episodic behavioral domains of every fluke specimen, all that combined into one complex multidomain system. Let us call it “the Fluke System”. Assuming, for instance, the ordinary Darwinian theory[8] we can easily identify the nature of mutual interactions between the two salient domains of the Fluke System. It is the evolutionary domain of the fluke species on the one side, and the domain of behavioral and life episodes of individual fluke specimens on the one side. The interactions can be roughly described as follows: the inherited properties shape individual behavioral patterns, and conversely, the successes and failures of behavioral episodes have a cumulative effect on the hereditary properties of the species (a more precise description would take into account the domain of individual life stories and the domain of populations).

Now we can claim that the ant-manipulating strategy is an essential property of the *whole* Fluke System.

Second example: The Expressive Language (revisited). Let us consider our second example of a natural language; call it "L". A completely different time scale applies (1) to L as a historically evolving diachronic entity, (2) to the learning process of speakers of L, and (3) to any particular act of uttering (or writing, listening or reading) a sentence in L. There is, nevertheless, a two-way interaction between the corresponding levels: for instance, each concrete utterance chooses words in L, obeys grammatical rules of L, and follows the habits prevailing at the respective historical moment among the speakers of L. On the other hand, the vocabulary, grammar, and habits of L evolve over long periods of time, and are subject to the accumulated influence of many actual utterances. Because of these interrelationships, we should realize that these different components are just different facets or manifestations of a single multidomain system. Let us call it "the Language System".

Similarly, as in the case of the fluke, the intelligent "inventions" of language, like the subjunctive mode in English, is not just a property of concrete speech episodes, nor of individual speakers, nor of the nonmaterial historical entity L, but rather a property of the Language System as a whole. The Language System interestingly differs from the Fluke System. Some of the domains of the former involve conscious intentional entities, namely the minds of the users who *intentionally* use the various features of L, like the subjunctive mode, and in this way *unintentionally* helps to preservation it in the language (I will return to it in Section 10.)

Third example: the Chess Machine (revisited). Our first idea may be to view a computer as a multilevel system with functional organization of the hierarchy of levels. For instance, classical computers comprised several clearly distinguishable hardware levels (from electronics to central processing units) as well as software levels (from machine code to programs in a high-level language). The conceivable complexity of such a system led some thinkers to an undue optimism about the possibility of the spontaneous emergence of mental phenomena in a hierarchically organized machine (cf. Hofstadter, 1979).

We should take into account, however, that contrary to our previous examples, the hierarchical organization of the computer is artificially constructed all the way down to the "silicon" level and also the types of interlevel connections are part of the prior design pro-

8. Here it is irrelevant whether the Darwinian theory yields a correct account of the fluke behavior or not.

ject. Thus, the designer(s) should be taken into account.

Let us consider "the Chess-Machine System" including tentatively the following most relevant causal domains: (1) the *physical domain* of the execution of the chess-playing program; (2) the program *performance domain* (described in the language of chess); and (3) the *programming domain* subsuming the relevant part of the mental domain of the programmer[9] while working on the chess-playing program. The first two domains are concerned with particular chess matches (actual or potential) whereas the programming domain involves the whole process of the development of the program. Thus, all concrete decisions made by the programmer (for instance, which particular utility function should be implemented) are part of the system as much as the physical processes in the computer.

What are the interdomain connections in this case? Here the situation is somewhat complicated by the involvement of the programmer's intentionality. In fact, it is the source of explicit intentional links from the programming domain to the performance domain. Moreover, it also establishes mutual connections between the physical and the performance domains, whereby certain objects in the former are assigned appropriate meanings in the latter. The linkage from the performance domain to the programming domain is obvious: the outcomes of actual or imagined chess matches may cause further development of the program.

So far so good, but *who* actually defeated Kasparov? It seems that analogously to the previous cases we should not isolate one of the domains and look for a "responsible" entity in there. Neither the computer alone defeated Kasparov (as a machine it could not do more than just to follow the laws of physics), nor the programmer (nobody expects him to be an excellent chess player). What remains is the Chess-Machine System as a whole.

Fourth example: The Rational Mind (revisited). The ant-manipulating strategy of the fluke, the subjunctive mode of a language and the winning chess program were just three concrete examples of phenomena in multidomain systems with rather surprising degree of sophistication. This may motivate us to wonder about the "system" that is sophisticated *par excellence*: "the brain and its mind".

Perhaps we would like to start (as many theoreticians do) with the brain and identify an appropriate hierarchy of levels in which each level could have its characteristic language of description, type of described phenomena, and its own causal relations (hopefully even causal laws). If the brain could be compared to the classical computer (as some believe), it would be easier: even if it is rather difficult within one view to embrace all functional levels in the computer, we are at least able discern these levels and specify the way they interact. The human mind affords us, however, a different story. Even if neuroscientists can describe very thoroughly some of the causal domains of the (human or animal) brain, these "known" domains are separated from each other by a large hiatus in knowledge. Thus, it is more our wish than a real possibility to imagine the functional organization of the brain in the form of an intelligible hierarchy of levels. Even more audacious is the wish of some materialists and emergentists that our genuine mind would occupy a certain sufficiently "high" level in the very same hierarchy.

I take both wishes with doubts. Even if we were able to identify organizationally and

9. Actually of a team of programmers. I am using the singular just for simplicity.

functionally "dense" (fitting with each other) collection of causal domains, it would not guarantee that the web of interdomain interactions would allow us to arrange them into a simple hierarchy. And even if such hierarchy existed, this would not imply that the mental level belonged to it.

What differentiate living organisms (and brains) from machines are the very existence, interplay and mutual interaction of a large number of different causal domains. Thus, we have again a multidomain entity, let us call it "the Brain-and-Mind System", that seems to comprise many different causal domains. The complexity of the system is not as much related to the number of domains as to the fact that they are densely "packed" within relatively few spatio-temporal scales (even if some of the domains involve highly parallel functioning of an exceedingly large number of active units). Both density and interaction are crucial features here. The density makes it difficult to study the domains individually and the interactions between domains may require fundamentally new scientific approaches.

Analogously as in the previous examples, we may view mental phenomena as though they are sustained by certain emergent properties of the whole Brain-and-Mind System. For this, however, we need to develop a concept of emergence more general than the ordinary one.

10. Second-Order Emergence and Rationality

The examples in the preceding chapter implied that most impressive cases of apparently rational behavior arise in complex multidomain systems – they deserve to be called "complex" in the sense that their existence and durability requires nontrivial interactions between various causal domains (levels). Usually such systems are not conceived as one single entity, perhaps due to the fact that the relevant domains (levels) are conceptually isolated and/or that they substantially differ in their characteristic spatio-temporal scales.

An exemplary situation is sketched in a schematic way in Figure 4. Roughly speaking the scheme illustrates the history of a kind of entity in domain A (to be called "upper level") in mutual interaction with a population of individual instances (episodes, occurrences, tokens) of the same kind in domain B (to be called "lower level"). Let us ignore possible direct interactions between individuals in domain B. It is conceivable that under a change of perspective the same scheme repeats downward, upward, or sideward forming a larger multidomain system.

In reference to our examples, the scheme in Figure 4 may correspond to two main domains of the Fluke System or two main domains of the Language System.

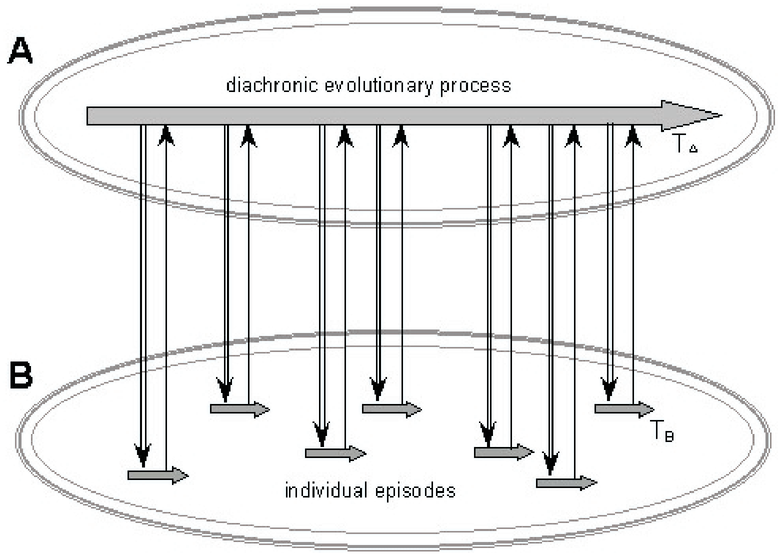

Figure 4. Two-level system $\mathbf{S}_{AB}$ evolving in domain A (in time scale T_A) with individual episodic events in domain B (events in typical time scale T_B). The upper-level process sets the parameters for episodic events (double arrows); the events contribute to the evolution of the process (single arrows).

For the attempt to look for a possible background of apparent rationality the usual concept of emergence (as a phenomenon at one level, supported or produced by events at another level) turns out to be insufficient. Elsewhere I proposed a new conception of emergence, the *second-order emergence* (Havel, 1993). Roughly speaking, an entity is second- order emergent if it arises from global interaction or "cooperation" of many domains of a complex multidomain system.

I am suggesting here that the term *emergent*, when applied to rationality, (or to purposiveness, inentionality, etc.), should imply the second-order emergence. This does not rule out attributing emergent rationality to endemic entities in a particular causal domain (that is, in a domain-specific discourse). However, then it expresses (unlike the "as-if" ascription, which suggests that there is "nothing behind" an appearance) our understanding that *behind* the appearance there is something more: a multidomain system with nontrivial interdomain interactions.[10] Thus it is, in effect, much stronger attribution of rationality than a mere "as-if" attribution. On the other hand, emergent rationality may be a weaker phenomenon than intrinsic rationality (in Searle's sense) because no claim is made about the necessity of involvement of a conscious self associated with the domain in question.

We may demonstrate our conception of emergent rationality with the Fluke System, for simplicity reduced to two interacting levels as in Figure 4. As observers, we would ad-

10. An alternative formulation may associate emergent rationality solely with the whole multidomain system and in the domain-specific discourse only use the term figuratively.

mire the apparently rational ant-manipulating behavior at the specimen level (domain B). But there it is (according to current views) just a fixed genetically preprogrammed behavior. At the same time, we know that the program is a result of much slower upper-level evolutionary process (in domain A). This process, in turn, could have only made the "decision" about the proper ant-manipulating strategy after a long history of testing various alternative strategies with individuals in domain B. Thus the apparent rationality is the outcome of "cooperation" of both domains significantly separated in time scales.

The example hints at the possibility of the decision being made in one domain while the reasons justifying it belong to quite a different domain. This possibility opens an interesting area of investigation: *How to understand a multidomain system that exhibits apparent rationality in one of its domains while it involves conscious and intentional agencies in quite another domain?*

One category of such systems, exemplified by the Language System, includes various social institutions and organizations: legal systems, political structures, corporations, science, art, games, etc. They have the common property that their overall history (in a larger time scale) depends in a certain known or unknown way on a multitude of episodic events or acts controlled by conscious agencies at a "lower" level (in a smaller time scale). These agencies have specific intentions and goals and with the freedom of choice.

Consider again a two-level system in Figure 4. Assume that the episodic events are at the lower-level (domain B) and that the upper-level evolving process (in domain A) is neither random nor under anybody's direct control. Rather it is dependent on the cumulative effect of the lower-level episodic actions made by intentional agencies. In a special case, when these agencies are unaware of the dynamics at the upper level, the whole system has similar non-personal character as many other systems encountered in nature. This may be the case of the Language System, if we assume that the speakers, who intentionally use various features of the language, are unaware of possible reverse effects of their speech acts on the language as such.

Now assume that the agencies at the lower level, besides being consciously concerned with individual episodic events, also know about the existence and dynamics of the upper-level global process. Then several cases can be discerned. First case: the lower-level agencies have no desires to influence the upper-level process (they "think and act locally"); second case: they have such desires but they have no idea how to realize them; third case: they believe they know how to realize such desires and they behave accordingly ("think globally, act locally"). Of course, they may be mistaken in their beliefs – which happens to be a frequent case in human societies.

In the last case, it is conceivable that individual decisions at the lower level collectively (and successfully) favor some development at the upper level (on which the agencies may agree, perhaps only by a majority). We might want to talk in such a case about *emergent collective rationality.* Even if it is not based on any upper-level consciousness, in the minds of the lower-level agencies (now in the role of observers) the upper-level dynamics may well produce a (false) impression that it is under control of a virtual upper-level rational agency. The impression may be even supported by the fact that the influence of each lower-level individual act on the behavior of the upper-level process may be infinitesimal due to the enormous number of other lower-level individual acts that jointly participate in affecting the direction of the overall process.

Let us turn next to another category of multidomain systems exemplified by our second two examples, the Chess-Machine System and the Brain-and-Mind System.

For the first case, consider again the question "*Who* defeated Kasparov?" in reference to the concrete May 1997 match (to avoid discussion of the game of chess in the large, or the field of AI in the large: this would bring us back to the previous category).

In this case it is popular to attribute the achievement – and, in general, the competence of rational decision making – either to the "Deep Blue" machine alone, or to the programmer (programming team) alone. However, in the framework suggested in this study, we should consider the whole multidomain system it its entirety, including the physical domain, the performance domain and the programming domain (perhaps, some other decompositions may be appropriate). The apparent rational behavior in the performance domain can be then grasped as a second-order emergent property of the whole Chess-Machine System.

The issue of machine "mentality" has been extensively studied since the ascent of programmed computers an it is not my aim here to sort and review various diverse views. It seems natural and compatible with our approach to talk in this case about *derivative rationality* in a similar sense as Haugeland (1998) talks about derivative intentionality. The person who creates the program is well aware of the ends of his actions (the programming domain includes the relevant part of his mental domain) whereby these ends are formulated in the language of the performance domain (e.g., what amounts to be a winning strategy). Thus, the apparent rationality in performance of the machine is derived from the rationality of the programmer.

Finally, let me shortly mention the Brain-and-Mind case that is, indeed, the most interesting one, even if it is not, in fact, quite in the scope of this study in which we mostly deal with other than intrinsic rationality. What can make it attractive for us is that with the mental domain both the intrinsic rationality (involving consciousness) and the apparent rationality are associated, while the brain domain (or a variety of biological-physical domains related to the functioning of the brain) is just a structure composed of apparently non-rational elements (neurons and their collectives).

In view of the cases mentioned earlier, we may pose a question to what extent the concept of emergent rationality, based on the idea of second-order emergence phenomena in complex multidomain systems, may help us to understand also the Brain-and-Mind System. If we believed, like earlier Searle, that our (individual, human) mental phenomena are caused and sustained by "blind" neurophysiological processes in the structure of the brain, why should we be so reluctant to ascribe analogous phenomena, intrinsic rationality included, to some higher structures (languages, human organizations, etc.) based on, or composed of, an immense number of mutually interacting (even possibly intentional and rational) lower-level individuals?

Or conversely: assume that neurons (or other basic units) in the brain are all conscious, intentional and rational little creatures. What difference would it make if these creatures, in addition to their local interests, were aware of the existence of upper-level phenomena and deliberately influenced them? (Let me parenthetically remark that such a question may become rather relevant in the future Internet society.)*

* The research is partly supported by the institutional research project MSM 110000001 of the

Literature

Chalmers, D. J. 1996 *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Dawkins, R. 1982 *The Extended Phenotype*, Oxford: W. H. Freeman, 1982.

Haugeland, J. 1998 *Having Thought: Essays in the Metaphysics of Mind*, Cambridge, Mass.: Harvard University Press.

Havel, I. M. 1993 "Artificial Thought and Emergent Mind", in *Proceedings of the International Joint Conference on Artificial Intelligence '93*. Denver, CO: Morgan Kaufman Professional Book Center, 758-766.

Havel, I. M. 1996 "Scale Dimensions in Nature", *Int. Journal of General Systems*, 24, 295–324.

Hofstadter, D. R. 1979 *Gödel, Escher, Bach: An Eternal Golden Braid*, New York: Basic Books.

Kim, J. 1974/1993 "Noncausal connections", *Noûs*, 8, 41-52, also in: Kim, J. 1993 *Supervenience and Mind. Selected philosophical essays*, Cambridge: Cambridge University Press, 22-32.

Kim, J. 2000 *Mind in a Physical World: Ana Essay on the Mind-Body Problem and Mental Causation*, Cambridge, Mass.: The MIT Press.

Moravec, H. 1999 *Robot: Mere Machine to Transcendent Mind*, Oxford: Oxford University Press.

Nagel, E. 1961 *The Structure of Science: Problems in the Logic of Scientific Explanation*, New York: Harcourt, Brace & World.

Oaksford, M. and Chater, N. 1998 *Rationality in an Uncertain World* (Essays on the Cognitive Science of Human Reasoning), Psychology Press.

Scott, A. 1995 *Stairway to the Mi*

nd, New York: Springer-Verlag.

Searle, J. R. 1983 *Intentionality*, Cambridge: Cambridge University Press.

Searle, J. R. 1992 *The Rediscovery of the Mind*, Cambridge, Mass.: The MIT Press.

Searle, J. R. 2000 *Rationality in Action*, Cambridge, Mass.: The MIT Press.

Varela, F. J., Thompson, E., and Rosch, E. 1991 *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, Mass: The MIT Press.

Czech Ministry of Education. I owe a special debt to my colleagues in CTS for interesting conversations about the subject. Thanks are also due to Kevin Coffey for correcting my English in sections that the reader may easily recognize.