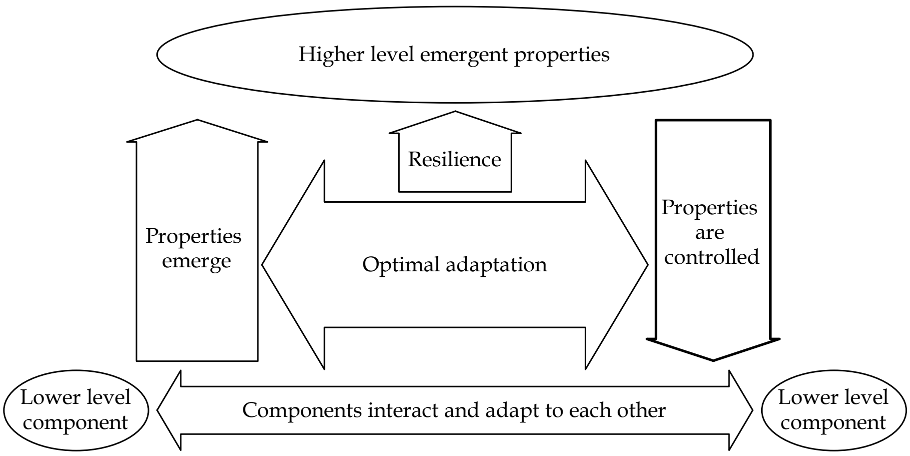
NLR Air Transport Safety Institute
Research & Consultancy

Executive summary

An investigation of emergent behaviour viewpoints in literature, and their usability in Air Traffic Management



Problem area
There is wide consensus that it is essential for future Air Traffic Management (ATM) developments to get a grip on *emergent behaviour*, i.e. on behaviour that cannot be directly contributed to individual components, but that somehow 'emerges' anyway. With the introduction of advanced ATM concepts as considered in the SESAR program (Single European Sky ATM Research), new behaviour and hazards will emerge that have not yet been seen before. In addition to emergent hazards, also positive emergent behaviour is of importance; unique qualities can be attributed to systems that show emergent behaviour, such as robustness, resilience, and the ability to find a reasonable solution quickly without complete knowledge or understanding.

By now, there is an abundance of literature on the topic 'emergence'. Some papers aim at understanding the mechanisms behind emergent behaviour; other papers try to formulate proper definitions; yet other papers argue whether or not emergence actually exists, or pose criticism to different visions. When put together, the literature proposes numerous

Report no.
NLR-TP-2011-444

Author(s)
M.H.C. Everdij
J.J. Scholte
H.A.P. Blom
S.H. Stroeve

Report classification
UNCLASSIFIED

Date
July 2011

Knowledge area(s)
Vliegveiligheid (safety & security)

Descriptor(s)
Emergence
Literature study
Air traffic management
Interactions
Emergent behaviour

This paper has been presented at the First Complex World Annual Conference, Seville, Spain, 6-8 July 2011.

NLR Air Transport Safety Institute
Research & Consultancy

definitions of and visions on emergent behaviour, or characterizations and itemizations of the properties of emergent phenomena, such that an accepted unambiguous definition for the concept of emergence does not exist.

Description of work
The approach taken in this paper is to identify the main emergent behaviour viewpoints from literature, and to connect these to practical examples from ATM. The eventual objective is not to try to formulate one all-encompassing definition, but to consider each vision as another useful perspective that provides some light and understanding on this phenomenon. This should lead to a broad better understanding of the term emergence, the potential mechanisms behind it, and the potential application to ATM.

Results and conclusions
Even though all viewpoints outlined in this paper have their merits and provide additional insight into emergent phenomena, some viewpoints appear to be easier connected to ATM examples than others.
We identified ATM examples for the following viewpoints:

- Emergence as a special case of 'synergy',
- Emergence as tension between 'reductionism' and 'weak downward causation',
- Emergence as tension between 'reductionism' and 'autonomy',
- 'Weak emergence by simulation',
- 'Weak emergence as surprise',
- Emergence by 'adaptation'.

For other viewpoints, we were not able to find ATM examples, either due to the absence of an unambiguous definition, or due to the observation that there are not many examples for this viewpoint beyond ATM either:

- 'Strong downward causation',
- 'Supervenience',
- 'Strong emergence'.

Finding examples for these remaining viewpoints is left for further work.

The overview of the main viewpoints and characterisations of emergence that appear in the literature, and the illustration that many of these characterisations can be observed in air transport safety as well as air transport risk, provides an alternative perspective of looking at air transport safety related phenomena, with the aim to improve our understanding of their mechanisms.

At a next step, the insight may help to identify new methods or approaches to identify or analyse emergent behaviour in air transport.

NLR Air Transport Safety Institute

Anthony Fokkerweg 2, 1059 CM Amsterdam,
P.O. Box 90502, 1006 BM Amsterdam, The Netherlands
Telephone +31 20 511 35 00, Fax +31 20 511 32 10, Web site: http://www.nlr-atsi.nl

NLR Air Transport Safety Institute
Research & Consultancy

NLR-TP-2011-444

An investigation of emergent behaviour viewpoints in literature, and their usability in Air Traffic Management

M.H.C. Everdij
J.J. Scholte
H.A.P. Blom
S.H. Stroeve

This paper has been presented at the First Complex World Annual Conference, Seville, Spain, 6-8 July 2011.

Contract number	4003131
Division	Air Transport
Distribution	Unlimited
Classification of title	Unclassified
	July 2011

Approved by:

Author	Reviewer	Managing department
13/10/11 21/10/11 14/10/11 14/10/11	Conference	5/11/11

ABSTRACT

There is an abundance of literature on the subject of emergence. One mainstream uses the term to describe the way complex patterns arise out of a multiplicity of relatively simple interactions, for example the literature on emergent behaviour in socio-technical systems. Another mainstream, in philosophy, aims at properly defining and characterising emergence. For development of safe future ATM, all such areas are of importance. There is sound reason to exploit the power of multi-agent based modelling and simulation in the design of advanced ATM, just as it is common practice to do so in other complex industries. At the same time it also is clear that the performance of future ATM will critically depend upon their humans and organizations, and this forms a good rationale for considering a wider scope than what is needed for a multi-agent based approach. The best way to handle this for safe future ATM is that we learn about similarities and differences between the existing emergent behaviour views, and their applicability to ATM. The aim of this paper is to make a start with this learning.

Contents

Abbreviations

ATM	Air Traffic Management
ICAO	International Civil Aviation Organization
R/T	Radio Telephony
SESAR	Single European Sky ATM Research

This page is intentionally left blank.

1 Introduction

There is wide consensus (e.g., [Shah et al, 2005], [SESAR, 2008], [Eurocontrol, 2010]) that it is essential for future Air Traffic Management (ATM) developments to get a grip on emergent behaviour. [SESAR, 2008] explains that with the introduction of advanced ATM concepts as considered in the SESAR program (Single European Sky ATM Research) yet unknown emergent risk may appear: new behaviour and hazards will emerge that have not yet been seen before. In addition to emergent hazards, also positive emergent behaviour is of importance; unique qualities can be attributed to systems that show emergent behaviour (i.e., emergent systems), such as robustness, resilience, and the ability to find a reasonable solution quickly without complete knowledge or understanding [Beart, 2003].

Good historical overviews of the origin of the term emergence are given by [Goldstein, 1999] and [Corning, 2002]. Aristotle [Aristotle, 350 BC] already referred to the notion as "the whole is something over and above its parts, and not just the sum of them all…". The term "emergent" is said to be coined 135 years ago by the English philosopher G.H. Lewes [Lewes, 1875], who defined emergents in the context of chemical compounds in contrast to resultants:

> "Every resultant is either a sum or a difference of the co-operant forces; their sum, when their directions are the same – their difference, when their directions are contrary. Further, every resultant is clearly traceable in its components, because these are homogeneous and commensurable. It is otherwise with *emergents*, when, instead of adding measurable motion to measurable motion, or things of one kind to other individuals of their kind, there is a co-operation of things of *unlike kinds*. The emergent is unlike its components insofar as these are incommensurable, and it cannot be reduced to their sum or their difference."

This view of emergence amounts to saying that the properties of a "whole" (i.e. *systemic qualities*) cannot be deduced by summing or averaging the properties of its components. One may see that, compared to the notion of Aristoteles, Lewes argues that for emergence to occur, there is co-operation between components of unlike kind.

In the 1920s and 1930s, the study of emergence was low, but it found renewed interest (it 're-emerged') in the last decades of the twentieth century, with the growth of scientific interest in the phenomenon of complexity and the development of new non-linear mathematical tools, [Corning, 2002]. According to [Goldspink & Kay, 2009], the concept of emergence first found wide adoption within the philosophy of science, but it more recently has been advanced within three distinct streams: *philosophy*, particularly of science and mind; systems theory, in particular complex systems; and *social science*, where it has largely been referred to under the heading of the micro-macro link and/or the problem of structure and agency. [Goldspink & Kay, 2009] also note that there has been relatively little cross fertilization of thinking between these streams.

By now, there is an abundance of literature on the topic. Some papers aim at understanding the mechanisms behind emergent behaviour; other papers try to formulate proper definitions; yet other papers argue whether or not emergence actually exists, or pose criticism to different visions. When put together, the literature proposes numerous definitions of and visions on emergent behaviour (e.g., [Chalmers, 2002a], [Hoyningen & Lohse, 2009]), or characterizations and itemizations of the properties of emergent phenomena (e.g., [Goldstein, 1999], [Goldspink & Kay, 2009], and [Peterson, 2006]), such that an accepted unambiguous definition for the concept of emergence does not exist. The approach taken in this paper is to identify the main emergent behaviour viewpoints from literature, and to connect these to practical *examples* from ATM. The eventual objective is not to try to formulate one all-encompassing definition, but to consider each vision as another useful perspective that provides some light and understanding on this phenomenon. This should lead to a broad better understanding of the term emergence, the potential mechanisms behind it, and the potential application to ATM.

The following Sections 2 through 8 each explain a main viewpoint of emergent behaviour from literature. Per viewpoint, an example from ATM is provided. Section 9 provides concluding remarks.

2 Emergence as a special case of Synergy

[Corning, 2002] argues that the term emergence is ill-defined and confusing. Therefore, he prefers to define emergence as a special case of the better understood concept of *synergy*, which he defines as follows:

> "Synergy refers to the combined (cooperative) effects that are produced by two or more particles, elements, parts or organisms – effects that are not otherwise attainable".

Hence, a 'whole' only produces synergy if it shows an effect that is not otherwise attainable, and synergy is not 'more' than the sum of the parts, just different. Corning adds that synergetic effects are not vague but are, as a rule, very concrete and measurable. One can test for the presence of synergy by removing one or more important parts and observing the consequences. Different types of synergy are identified, such as:

- Functional complementarity effects produced by new combinations of different parts. An example is table salt: the parts sodium and chloride are toxic, but their combination no longer is.
- Division of labour, in which processes that are difficult to combine for one actor are conducted by different actors together.
- Synergy of scale, i.e. an aggregation of interchangeable, like-kind parts that produce unique cooperative effects. An example is a river or a pile of sand.
- Threshold effects of synergy of scale, such as a river becoming a flood.
- Cost and risk sharing.
- Information sharing and joint decision-making.

Corning next defines *emergence* as a subclass of synergy as follows:

> "Emergence would be confined to those synergistic wholes that are composed of things of "unlike kind". It would also be limited to "qualitative novelties"— i.e., unique synergistic effects that are generated by functional complementarities, or a combination of labour."

Emergent effects would be associated specifically with contexts in which constituent parts *with different properties* are modified, re-shaped or transformed by their participation in the whole. Hence, water and table salt would be considered emergent phenomena, but a sand pile or a river would not be. Note that this is in agreement with Lewes' definition of emergence [Lewes, 1875], see Section 1, who used the term "of unlike kind" as well.

ATM example

Many different humans and technical systems are involved in ATM operations: multiple air traffic controllers, pilots in multiple aircraft, technical systems such as aircraft systems, communication, navigation, and surveillance systems. All these components are clearly of ‘unlike kind’, and co-operate according to established procedures and often unwritten ways of working, aiming at a safe and efficient operation. We can do the Synergy test: if we leave out an important part such as the pilots, we would obviously have a different operation. Also different types of synergy can be identified: the pilots and the air traffic controller have different complementary roles to play (division of labour), and communicate vital decision information via R/T (information sharing). Finally, qualitative novelties in the sense of synergetic effects generated by functional complementarities, or combination of labour can be recognized: the controllers and pilots for example have functional complementarities which together ensure that aircraft fly in safe and orderly flows.

3 Emergence as tension between Reductionism and Downward causation

Several authors have looked further into the relationship between the 'components' of an emergent on the one hand, and the 'whole' on the other hand. The resulting visions on emergence can be characterized as a tension between two philosophical extremes: Reductionism, and Downward Causation.

- *Reductionism* (also called *Upward Causation*) argues that a complex system is nothing but the sum of its parts, and that an account of it can be reduced to accounts of individual constituents. Reductionism does not preclude the existence of what might be called emergent phenomena, but it does imply the ability to understand those phenomena completely in terms of the processes from which they are composed. This reductionist understanding is very different from that usually implied by the term 'emergence', which typically intends that what emerges is more than the sum of the processes from which it emerges [Lewes, 1875]. [Corning, 2002] notices a shift in the meaning of reductionism: in the 19th and 20th century, it meant an understanding of the 'parts' of the system; modern-day reductionists, by contrast, speak of the parts *and* their interactions. [Corning, 2002] considers the latter formulation not to be proper reductionism, but rather 'system's science in disguise', since the interactions *are* the system.
- *Downward causation* may be seen as the opposite of the reductionism perspective, but several interpretations of the term are in circulation. A 'strong' definition is proposed by [Sperry, 1964], who stated that in downward causation the higher level laws have power to *downwardly control* the lower level laws. [Chalmers, 2002] and [Bedau, 2002] consider this view not very scientifically useful; [Chalmers, 2002] identified only one example of such strong downward causal phenomenon, which is *Conscience*. For this and other reasons, other authors have started to propose 'weaker' definitions. [Campbell, 1974] defined downward causation as: 'all processes at the lower level of a hierarchy are *restrained by* and act in conformity to the laws of the higher level'. For this weaker definition, many more examples can be identified, see e.g. [Heylighen, 1995]. In addition, the weaker definition leaves more freedom to be combined with reductionism views.

From the weaker definitions of downward causation, it may be noted that many phenomena considered as emerging contain aspects of both reductionism and downward causation. [Heylighen, 1995] for example combines these concepts by noting that the whole is to some degree constrained by or determined by the parts, but at the same time the parts are to some degree constrained by or determined by the whole. The difference is that in downward causation, determination is ‘not complete’. This makes it possible to formulate a clear systemic stance, without lapsing into either the extremes of reductionism or of holism. [Rockwell, 2002] argues that emergent causal properties are necessary for downward causation, though not sufficient. [Koestler, 1969] uses the metaphor of Janus (i.e. the ancient Roman god with two faces in opposite directions) to illustrate how the two perspectives (holistic vs. reductionist) should be treated as perspectives, not exclusives, and should work together to address the issues of *emergence*. Further, he notes that the ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. Such constructionist hypothesis breaks down when confronted with the difficulties of scale and complexity. At each level of complexity, entirely new properties appear. [Corning, 2002] argues that since there are various levels of hierarchy, emergence also occurs at various levels. At each level there is downward causation as well as upward causation. In addition, causation is iterative: synergistic effects produced by emergent systems are also causes. This is called the Synergistic Hypothesis. [Davies, 2006] argues that the mechanism of downward causation can usefully be considered in terms of boundaries. Novelty of the whole, he argues, “may have its origin in a system being ‘open’; if novel order emerges, it must do so within the constraints of physics”. But he concludes, “openness to the environment merely explains why there may be room for top-down causation; it tells us nothing about how that causation works.”

In the literature, e.g. [Castellano, 2010], there are also several discussions regarding *problems* with downward causation. [Bedau, 2002] summarises the three main problems as follows:

- The very idea of downward causation is incoherent, since if higher level properties are emerging from lower level properties, how can they also causally influence these lower level properties?
- Even if downward causation would be coherent, it would make a difference only if it would violate micro causal laws.
- Even if downward causation would be coherent and consistent with fundamental micro laws, then any macro-level cause that has micro-level effect (i.e., any downward causation), will compete for explanatory relevance

with the micro-level explanation; the micro-level explanation would then be considered as more fundamental.

[Kim, 2006] discusses two fundamental unresolved issues for emergence. The first is that of giving a ‘positive’ characterization of emergence; the second is to give a coherent explanation of how ‘downward’ causation is able to avoid the problem of overdetermination. [Brown, 2010], however, criticizes Kim’s discussion, by explaining that Kim’s concept of emergence is based on the notions of emergence as known in the early 20th century.

<u>ATM example</u>

[Chalmers, 2002] identified “conscience” as ‘the only’ example of the ‘strong’ definition for downward causation, which indicates that attempts to identify further examples in ATM are likely to be futile.

An ATM example of the ‘weaker’ definition for downward causation in combination with reductionism is safety regulation. In the early days of air traffic, the number of aircraft flying was very limited, and there was not much need for ATM safety regulation. As the volume of air traffic increased, established working processes became harmonised, standardised, and captured in regulations (reductionism). Next, these regulations had to be complied to, and in turn influenced the working processes (downward causation).

One may look at the occurrence of major accidents as another example. The unfortunate or erroneous interaction of several actors and components may lead to an accident (reductionism), while after such accident appropriate measures are often taken to reduce the possibility of similar accidents occurring in the future (downward causation).

This page is intentionally left blank.

4 Emergence as tension between Reductionism and Autonomy

A similar field of tension between two 'extreme' views is considered by, e.g., [Bedau, 2002] and [Wyss, 2004]. [Bedau, 2002] introduces two 'hallmarks' of how macro-level emergent phenomena are related to their micro-level bases:

- Emergent phenomena are autonomous from underlying processes.
- Emergent phenomena are dependent on underlying processes.

He explains that any way of simultaneously meeting both these hallmarks (autonomy vs. reductionism) is a candidate notion of emergence. He notes that these two hallmarks are vague, but provide structure and a framework for comparing the various notions.

[Wyss, 2004] considers these same two hallmarks, which he refers to as 'central ideas', in the context of explanation of the existence of organisms. He argues that different current definitions of emergence are mixtures of these two 'central ideas':

- The distinctiveness of emergent properties: Life and mind are essentially distinctive from physical matter;
- The dependence of emergent properties: Life and mind are dependent on physical matter.

The first idea (autonomy) gives a new identity to an emerging property. The second idea (reductionism) is that ultimately, everything is made of its parts.

Downward causation may be argued to follow from the first 'hallmark' or 'central idea' (i.e. autonomy), in the view that "nothing is real unless it has *causal powers*" ([Wyss, 2004], quoting [Alexander, 1920]). This implies causal powers that are *irreducible* and *fundamental*; otherwise the emergent property would not be autonomous. As a consequence, "the causality of emergent properties is inexplicable in terms of, and theoretically unpredictable from, those of their base properties".

<u>ATM example</u>

Air transport safety may be viewed as an example that meets Bedau's two hallmarks or Wyss' two central ideas. Clearly, air transport safety is dependent on underlying processes as the working procedures of individual components of the air transport system. Also, air transport safety can be considered as autonomous, or at least as irreducible and fundamental: safety is a result of interactions between all elements of air transport, but somehow it cannot be attributed to one or multiple of these components themselves. Studies have argued that the ability of the human (pilots, air traffic controllers) to be creative, to improvise, and to co-ordinate in ways that are not in written procedures, makes air transport safe and resilient, but apparently it is still very difficult to see where safety really comes from (irreducibility).

5 Emergence and Supervenience

A further concept of relevance for views of emergence, is *Supervenience*. There are numerous heated discussions on the term and on how it should be interpreted and defined, and there is clearly no consensus. According to [Freewont, 2010]:

> “Supervenience is—or has become—too complicated to be given any kind of adequate account here. The lack of consensus among its proponents is striking. Its definitional vagaries, conceptual ambiguity, and appeal to a very diverse body of researchers has resulted in a chaotically incongruent body of theory.”

Some of the ‘definitions’ are, see e.g. [Chalmers, 1996], [Armstrong, 1997], [Kim, 1998], [Johansson, 2002], [Wyss, 2004], [Cooper, 2009]:

- For two families of properties P and Q, P supervenes on Q, if two things that are indiscernible with respect to Q are indiscernible with respect to P.
- An instance of P supervenes on an instance of Q, if there can be no change in P without corresponding change in Q (though there may be Q-change without P-change).
- P-properties supervene on Q-properties if and only if anything that has a P-property has some Q-property such that anything that has that Q-property also has that P-property.
- Entity P supervenes upon entity Q if and only if it is impossible that Q should exist and P not exist, where Q is possible.
- P-properties supervene on Q-properties if no two possible situations are identical with respect to their Q-properties while differing in their P-properties.

Different forms of supervenience are defined in e.g. [Moyer, 2000], [Freewont, 2010], [Kim, 1984, 1993], [McLaughlin & Bennett, 2010], including ‘weak individual supervenience’, ‘regional supervenience’, ‘similarity-based supervenience’, ‘global supervenience’, ‘multiple domain supervenience’. The precise relationships between these different forms are still food for debate.

Discussions relating supervenience with *emergence* are primarily discussions related to emergent properties being supervenient on lower-level properties, see e.g. [Horgan, 1993]. There is also a relation between supervenience and reductionism, although the precise relation is yet unclear: According to

[McLaughlin & Bennett, 2010], “everyone agrees that reduction requires supervenience”. However whether supervenience is *sufficient* for reduction is an open question.

<u>ATM example:</u>
Given the diversity in definitions on supervenience, we decided to leave providing a clear and illustrative ATM example to future work.

6 STRONG AND WEAK EMERGENCE

An important step in dealing with the tension between reductionism and downward causation discussed in Section 3, is the definition of different forms of emergence. After initially distinguishing strong and weak emergence [Bedau, 1997], Bedau distinguishes three forms in [Bedau, 2002]:

- *Nominal emergence*, which is a notion of a "macro property" that cannot be a "micro property".
- *Strong emergence*, which is nominal emergence in which the emergent properties are supervenient properties with irreducible downward causal powers.
- *Weak emergence*, which is nominal emergence that not strong. The system's global behaviour derives from the operation of micro-level processes, but the micro-level interactions are interwoven in such a complicated network that the global behaviour has no simple explanation.

Apparently, nominal emergence can be sub-divided into the other two forms.

Bedau argues that 'strong' emergence has had a prominent place in the philosophical discussions but that its scientific credentials are very poor, whereas 'weak emergence' is consistent with materialism and scientifically useful:

> "Although strong emergence is logically possible, it is uncomfortably like magic. How does an irreducible but supervenient downward causal power arise, since by definition it cannot be due to the aggregation of the micro-level potentialities? Such causal powers would be quite unlike anything within our scientific ken. This not only indicates how they will discomfort reasonable forms of materialism. Their mysteriousness will only heighten the traditional worry that emergence entails illegitimately getting something from nothing." [Bedau, 1997]

Since in Weakly emergent phenomena there is no downward causation, these phenomena can be derived from full knowledge of the micro-level properties, *but only in a certain way*. Bedau proceeds to defend one version of weak emergence (noting that there are other versions as well), which is as follows:

> "A nominally emergent property of a locally reducible system is called *weakly emergent* if it is derivable from all of the micro facts of this system, but *only by simulation*."

Here, in a simulation, each derivation iterates step by step through the aggregation of local interactions among the micro-elements. Bedau also notes that there are various degrees of weak emergence, e.g. properties that are derivable without simulation in principle, but in practice must be simulated; properties that are underivable except by finite feasible simulation; properties that are underivable except by simulation, but the requisite simulation is unfeasible or infinite. He also notes a distinction between *thinking* that some phenomenon is weakly emergent compared to some phenomenon being weakly emergent: one may believe that a phenomenon can only be obtained by simulation, but fail to see that there is also a yet undetected short-cut explanation for it. Weak emergence is also referred to as Computational Irreducibility, which is characteristic for complex systems and it explains why computer simulations are a necessary tool in their study.

Several further authors describe different orders of emergence in attempts to classify different types of behaviour perceived as emergent (e.g., [Gilbert, 2002], [Rasmussen et al, 1996], [Ellis, 2006], [Goldspink & Kay, 2008, 2009], and [Baas, 1994]).

ATM example

An example in air transport operations of Bedau's view of weak emergence as something derivable from individual components but only by simulation is described in [Stroeve et al., 2011]. A concept of operations is considered in which frequent active runway crossings take place on a departure runway in good visibility conditions. To limit potential risks related to such operation, the concept included a runway incursion alerting system to warn the air traffic controller in situations in which a departing and a crossing aircraft simultaneously make use or start making use of the runway.
According to early safety assessments using traditional approaches as fault trees and event trees, the alerting system would provide a significant risk reducing effect. However, according to Monte Carlo simulations of a dynamic risk model of the actors, systems and interactions, the risk decreasing contribution of the alerting system and the air traffic controller in the same concept appeared small. The key new insight obtained from the simulations was that in most situations in which the alerting system enables the air traffic controller to warn the pilot, the pilots of one of the involved aircraft has already identified and started to solve the conflict theirselves. If in time-critical situations the pilots did not detect the conflict, then it would often neither be resolved via the alerting system, e.g., because of a late alert, delay in the communication line between controller and pilots, or a late or inappropriate reaction of the controller or pilot.
The described effect was discovered only after developing and simulating a dynamic risk model that covered the totality of interactions of components

including their variability in performance over time. The complexity of air transport operations involves a combinatorial explosion of the many events that may occur in a dynamic way and the many involved uncertainties, such that certain aspects of safety risk can only be studied through simulation. The human mind is simply not able to grasp the many combinations of events occurring later or earlier than average, or resolutions of situations that are implemented in another way, not even when supported by graphical tools such as tree-based schemes or analytical equations. The Monte Carlo simulations made it possible to identify how the operation evolves through time in a dynamic way, addressing to a larger extent the combinatorial explosion and allowing specific behaviour to emerge.

This page is intentionally left blank.

7 EMERGENCE BY SURPRISE

Chalmers [Chalmers, 2002a] includes a notion of “unexpectedness” or “surprise” to the definition of emergence when providing alternative definitions for strong and weak emergence:

- A high-level phenomenon is *strongly emergent* with respect to a low-level domain when truths concerning the phenomenon are not deducible even in principle from truths in the low-level domain.
- A high-level phenomenon is *weakly emergent* with respect to a low-level domain when truths concerning the phenomenon are *unexpected* given the principles governing the low-level domain.

Note that this definition of strong emergence relates to ‘autonomy’, see Section 4.

Other authors also refer to the notion of surprise, such as [Sanz, 2004] who defines emergence as:

> “just systemic behaviour — nothing more, nothing less— that is difficult to predict in advance”.

Bedau [Bedau, 2002], explains that he left the notion of surprise absent on purpose, due to it being rather subjective. Bedau instead claims that, with his definition of weak emergence in terms of simulation (see Section 6), he is presenting objectivist approaches to emergence, though notes that his classification is not exhaustive.

[Johnson, 2006] seems to agree with Bedau on this by noting that problems may arise when engineers combine different meanings, such as a ‘surprise factor’ implicit in predictive approaches, while talking about the design of emergent properties. The predictive approach to emergence raises questions about the perspective of the person making the predictions; no-one has perfect knowledge. Johnson also argues that several accidents are attributed to emergent behaviour, even though often engineers have issued warnings about possible accidents that have been ignored until after an adverse incident has occurred (making the accidents less of a surprise to these engineers). Therefore, another, perhaps less subjective approach to the predictive issue is to define emergent phenomena as:

> “those that possess interesting properties that were *not included in the goals of the designer*.”

[Johnson, 2006] notes that the unpredictiveness notion of emergence can be contrasted with the notions of weak and strong emergence. He notes that "contradiction arises because engineers freely move from predictive definitions in which emergence is equated to a surprise and definitions of strong emergence where higher-level patterns can be used as design templates." Johnson argues that greater care must be taken when using terms such as 'emergence'.

[Stephan, 2002] also considers this notion of surprise, formulated as 'unpredictability', and takes an extended approach to the notions of weak and strong emergence by adding the concept of *diachronics*, which deals with phenomena happening over a period of time. To this end, Stephan considers three notions:

- *Novelty*: In the course of evolution exemplications of 'genuine novelties' occur again and again. Already existing entities form new constellations that produce new structures which may constitute new entities with new properties and behaviours.
- *Irreducibility*: Stephan notes that the failure to keep two different kinds of irreducibility apart has muddled recent debate about emergence. These two different kinds of irreducibility, downward causation and unanalyzability of systemic properties, have quite different consequences:
  - If, on the one hand, a system's property is irreducible because of the irreducibility of the system's parts behaviour on which the property supervenes, we seem to have a case of *downward causation*. This kind of downward causation does not violate the principle of the causal closure of the physical domain.
  - If, on the other hand, a systemic property is irreducible because it is not exhaustively analyzable in terms of its causal role, downward causation is not implied. Rather, it is dubitable how unanalyzable properties might play any causal role at all.
- *Unpredictability*: Likewise, an emergent property can be unpredictable if it is instantiated by a given kind of structure in a given kind of system, and that structure is unpredictable, or if even when the structure is predictable, the emergence of the property is in itself unpredictable, because that property is irreducible.

These notions are next used in a scheme of different types of emergence:

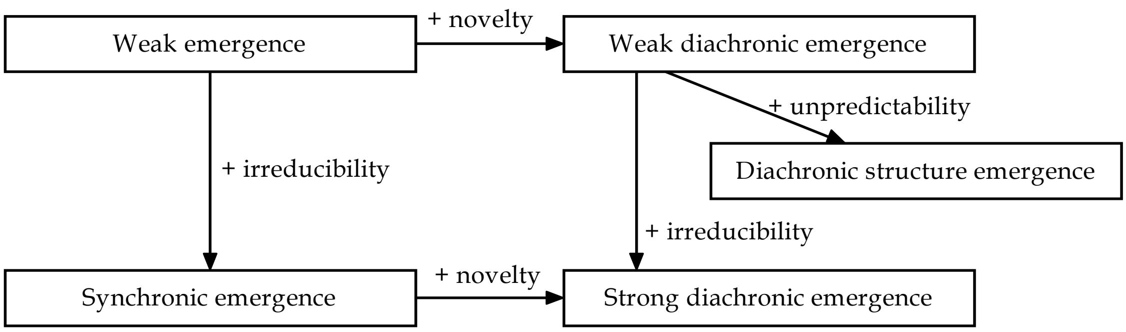


Figure 1: Different types of emergence are related by adding notions of novelty, unpredictability and irreducibility; figure based on [Stephan, 2002].

The scheme is explained as follows:

- *Weak diachronic emergence* results from weak emergence by adding a temporal dimension in the form of the thesis of *novelty*. Both versions are compatible with reductive physicalism.
- *Synchronic emergence* results from weak emergence by adding the thesis of *irreducibility*, i.e. being not reductively explainable. This version of emergence is not compatible with reductive physicalism. Synchronic emergence is of particular interest for the discussion of downward causation.
- *Strong diachronic emergence* only differs from synchronic emergence because of the temporal dimension in the thesis of novelty.
- In contrast, *structure emergence* is entirely independent of synchronic emergence. It results from weak diachronic emergence by adding the thesis of *structure-unpredictability*. Although structure emergence emphasizes the boundaries of prediction within physicalistic approaches, it is compatible with reductive physicalism, and so it is weaker than synchronic emergence.

<u>ATM example</u>

The surprise factor can also be identified for the example considered in Section 6. Initially, the expectation of the designers of the concept of operation was that the new runway incursion alerting system would make the controller significantly reduce the risk of the operation [Scholte et al., 2009]. A cause for this expectation is that air traffic controllers usually see aircraft brake after their halting instructions. In such case, the air traffic controller may perceive to have played a key role in resolving the conflict well, while not being aware that the conflict is actually solved by the pilots who had already identified and started to solve the conflict independently. It came thus as a surprise that the risk reducing effect of the alerting system was in practice much smaller.

This page is intentionally left blank.

8 Emergence by Adaptation

[Zarboutis & Wright, 2006] explain that in complexity science, each component interacts with its neighbouring ones by triggering the other at its border. Subsequently, this other component changes its internal organisation accordingly, through structural changes in order to assure the satisfaction of individual criteria, using local information and usually being unaware of the behaviour of the whole system. This process is known as adaptation. Subsequently, [Zarboutis & Wright, 2006] explain the notion of *emergence* as follows (see also the figure below):

> "The behaviour of the system is considered to be the product of local level interactions on various layers. On each of them, the components may locally interact; they can have their own structure and autonomous behaviour. However, only at a higher level, the properties of these interactions are evident. This operational mechanism of a complex system, where the product of local level interactions at a given level is evident at the higher one is called emergence and these higher level properties are called *emergent properties* of that level. This process of (bottom-up) emergence at the higher level undergoes simultaneously a form of (top-down) hierarchical control that wants to assure that the emergent properties are meaningful. The resilience of a complex system is dependent on a form of *optimal adaptation*, which goes through the *balance between emergence and hierarchical control*. The design of an effective hierarchical control system (e.g. a barrier system), so as to shape the emergent phenomena of the system, is the ultimate challenge for the engineering of complex systems."

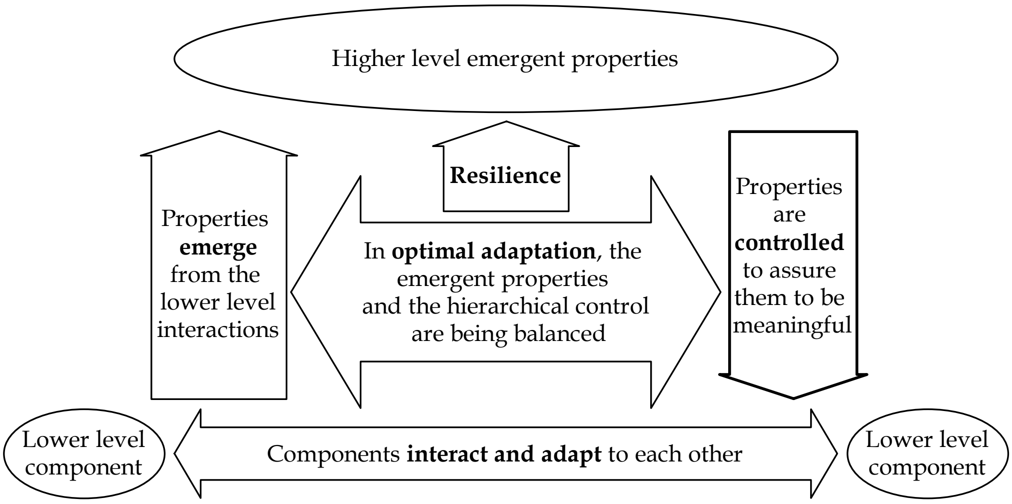


Figure 2: In Complexity Science, properties emerge from interactions between low level components and are being controlled in a balanced way to provide resilience. Figure based on theory proposed in [Zarboutis & Wright, 2006].

[Zarboutis & Wright, 2006] further explain that while strong predictions (i.e. who, when, where, etc.) of the behaviour of complex systems are impossible to achieve, it is possible to identify some recurrent patterns that, once emerged, divert the behaviour of a complex system towards a systemic collective event:

- *Self-Reference.* Is the pattern where an organisation produces by itself, the structure that creates itself, in a recurrent way in time, as a system evolves.
- *Infinite Loops.* Is a repetition of a set of processes until a condition is met. Once such condition is fulfilled then the loop exits to some given point. However, failing to meet this end condition, this iteration may continue endlessly and in most cases it can only be exited upon the imposition of some external force, intentional or unintentional.
- *Stigmergy.* Is an indirect mechanism of interaction where two parts each modify the environment, on which the other part adapts.

These primitive patterns need not be mutually exclusive; they usually stem out of lower level interactions and collectively give rise to some higher level patterns, which shape the system's emergence and which is referred to as co-adaptation.

Co-adaptation is responsible for some 'peculiar' collective phenomena, not immediately apprehensible through analytical reasoning. In general, co-adaptation is the process where two agents adapt to the same problem, each pursuing its own private goals. Within a complex organisation, co-adaptation is catalysed by stigmergy. Co-adaptation can have positive or negative effects on system performance. Under normal circumstances, the typical outcome of a co-adaptive act could lead to positive redundancy. But when the interacting agents pursue different or conflicting objectives, the system may collectively fail to adapt to external perturbations in a desired way and a systemic failure may take place.

The elimination of such patterns requires the removal of the sources that lead to their emergence. Thus, if we can assure for example that the interacting agents would always form common objectives, or that the necessary external forces would be present for an agent to exit the infinite loop that s/he is trapped into, then we will have achieved a more resilient organisation that would have the potential to create and maintain safety. Modelling the system through an explicit account of emergence, self-organisation and hierarchical control, the role of such patterns can become evident, while the relevant causes that diminish the resilience of the system can be identified.

ATM example
The air transport system of today is the result of decades of evolutionary development. Each day, it adapts to changing circumstances, such as changing passenger volumes, new rules, public safety perception, economic crises, environmental disruptions (e.g. volcano ash). There is a certain level of central control (e.g. ICAO), but most details are determined in a non-centralised way and adapted to the local circumstances (e.g. geographical location, weather circumstances, local passenger throughput), even though air transport is very international. Human operators are very well capable to adapt to the current circumstances. Higher level properties such as efficiency, capacity and safety emerge from the interactions and are balanced with each other and controlled, e.g. by flow management. Because of all these mechanisms, the air transport system appears to be very robust and resilient against disruptions and changes.

This page is intentionally left blank.

9 Concluding remarks

Air Transport Safety is the result of co-operation of multiple humans, organisations and technical systems, which are coupled in an indirect way, who work according to established procedures and many unwritten rules, and who interact with each other in a dynamic way in order to cope with the ever changing circumstances. As such, safety has many characteristics of an emergent phenomenon. This has a few negative aspects, such as a certain level of unpredictiveness, but it also has important benefits. The latter is shown by e.g. [Beart, 2003], who emphasises the positive aspects of emergence. He claims that there are things that emergent systems can do that other systems cannot:

- They are robust and resilient. There is no single-point of failure, so if a single unit fails, becomes lost or is stolen, the system still works.
- They are well-suited to the messy real world. Human-engineered systems may be 'optimal' but often require a lot of effort to design and are fragile in the face of changing conditions. Importantly, they don't need to have complete knowledge/understanding to achieve a goal (e.g. social systems in warehousing).
- They find a reasonable solution quickly and then optimise. In the real world, time matters - decisions need to be taken while they are still relevant. Traditional computer algorithms tend to not produce a useful result until they are complete (which may be too late, e.g. if you're trying to avoid an oncoming obstacle).

[Beart, 2003] further claims that these positive aspects are due to the individuals interacting with each other directly or indirectly (via their environment). Interacting via an effect on, and response to, their common environment is called *stigmergy* (see also Section 8).

In the philosophy, social sciences and systems theory literature, a lot has been written on emergence, but still there does not seem to be a well-established, unambiguous definition of the term. There are many discussions on what emergence is, where it may come from, and how the mechanisms may be understood. Different authors do not seem to agree, but many different viewpoints contribute to the further analysis and understanding of the phenomenon.

This paper has given an overview of the main viewpoints and characterisations of emergence that appear in the literature. Next, it has shown that many of these characterisations can be observed in air transport safety as well as air transport risk. This way, an alternative perspective was provided of looking at air transport safety related phenomena, with the aim to improve our understanding of their mechanisms. At a next step, the insight may help to identify new methods or approaches to identify or analyse emergent behaviour in air transport.

Further work

Even though all viewpoints outlined in this paper have their merits and provide additional insight into emergent phenomena, some viewpoints appear to be easier connected to ATM examples than others.

We identified ATM examples for the following viewpoints:

- Emergence as a special case of synergy (see Section 2),
- Emergence as tension between reductionism and weak downward causation (part of Section 3),
- Emergence as tension between reductionism and autonomy (see Section 4),
- Weak emergence by simulation (part of Section 6),
- Weak emergence as surprise (see Section 7),
- Emergence by adaptation (see Section 8).

Finding examples for the remaining viewpoints, i.e. 'Strong downward causation' (part of Section 3), 'Supervenience' (see Section 5), and 'Strong emergence' (part of Section 6) is left for further work. Note that for 'Strong downward causation' and 'Strong emergence', the non-ATM literature reports only one example, i.e. conscience, hence finding ATM examples for these viewpoints could be a challenge. For 'Supervenience', the challenge lies in first finding an unambiguous definition of the term itself.

Further insight can be gained in showing how the different viewpoints relate to each other, and how they can be used to explain the mechanisms behind emergence in air transport.

A further direction of possible future work is to improve the way that *safety analysis* addresses emergent behaviour in concepts of future ATM operations.

10 REFERENCES

[Alexander, 1920]	S. Alexander (1920). Space, Time, and Deity. The Gifford Lectures at Glasgow 1916-1918 (2 Vols.). London: Macmillan.
[Aristotle, 350 BC]	Aristotle, Metaphysics, 350 BC
[Armstrong, 1997]	Armstrong, D.M., 1997, A World of States of Affairs, Cambridge University Press, Cambridge.
[Baas, 1994]	N.A. Baas, Emergence, hierarchies, and hyperstructures, In C.G. Langton, ed., Artificial life III, Redwood city: Addison-Wesley, pp. 515-537. 1994
[Beart, 2003]	Pilgrim Beart, 2003, http://www.beart.org.uk/Emergent/
[Bedau, 1997]	Mark A. Bedau (1997). Weak emergence. In Tomberlin J (ed.), Philosophical Perspectives: Mind, Causation, and World, Vol 11, pp. 375-399, Blackwell, Malden (MA), USA, http://people.reed.edu/~mab/publications/papers/weak-emergence.pdf
[Bedau, 2002]	Mark A. Bedau, Downward Causation and the Autonomy of Weak Emergence, Draft for Principia, 2002, http://people.reed.edu/~mab/papers/principia.pdf
[Brown, 2010]	H. Brown, A critique of Jaegwon Kim's Emergence: Core Ideas and Issues, 27 November 2010, http://www.scribd.com/doc/44241899/A-Critique-of-Jaegwon-Kim-s-Emergence-Core-Ideas-and-Issues
[Campbell, 1974]	D.T. Campbell (1974): 'Downward causation' in Hierarchically Organized Biological Systems, in: Studies in the Philosophy of Biology, F.J. Ayala & T. Dobzhansky (ed.), (Macmillan Press), p. 179-186
[Castellano, 2010]	D.J. Castellano, Emergentism and Downward Causation, http://www.arcaneknowledge.org/science/vitapsych3.htm
[Chalmers, 2002]	D.J. Chalmers (2002), Strong and Weak Emergence, Republished in P. Clayton and P. Davies, eds. (2006) The Re-Emergence of Emergence. Oxford: Oxford University Press. http://consc.net/papers/emergence.pdf
[Chalmers, 2002a]	D.J. Chalmers (2002a), Varieties of Emergence, Templeton Foundation workshop on Emergence, Granada, August 2002, http://consc.net/papers/granada.html
[Cooper, 2009]	W. Cooper, Physicalism: Some Assumptions versus an Implication, Macquarie University, Sydney, Australia, 2009,

	http://www.arts.mq.edu.au/documents/NEO_Article_3_2009_Wilson_Cooper.pdf
[Corning, 2002]	Peter A. Corning (2002). The re-emergence of "emergence": A venerable concept in search of a theory. Complexity 7(6):18-30, http://www.complexsystems.org/publications/pdf/emergence3.pdf
[Davies, 2006]	P.C.W. Davies (2006), The Physics of Downward Causation, in P. Clayton and P. Davies (eds.), The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion, Oxford: Oxford University Press, pp. 35-51. http://www.ctnsstars.org/conferences/papers/The%20physics%20of%20downward%20causation.pdf
[Ellis, 2006]	G.F.R. Ellis (2006). On the Nature of Emergent Reality. In P. Clayton & P. Davies (Eds.), The Re-Emergence of Emergence: The Emergentist Hypothesis from Science to Religion. Oxford: Oxford University Press. http://www.mth.uct.ac.za/~ellis/emerge.doc
[Eurocontrol, 2010]	Eurocontrol (2010), European Operational Concept Validation Methodology, version 3.0, Volume II, available on http://www.eurocontrol.int/valfor.
[Freewont, 2010]	Freewont, Supervenient Downard Causation, 2010, http://www.freewont.org/2010/07/supervenience.html
[Gilbert, 2002]	N. Gilbert (2002). Varieties of Emergence. Paper presented at the Social Agents: Ecology, Exchange, and Evolution Conference Chicago. http://www.eicstes.org/EICSTES_PDF/PAPERS/Varieties%20of%20emergence%20(Gilbert).pdf
[Goldspink & Kay, 2008]	C. Goldspink & R. Kay (2008), Agent Cognitive capabilities and Orders of Emergence: critical thresholds relevant to the simulation of social behaviours. Proceedings of AISB Convention, Communication, Interaction and Social Intelligence Aberdeen, UK April 1-4, 2008. http://epubs.surrey.ac.uk/1603/1/fulltext.pdf
[Goldspink & Kay, 2009]	C. Goldspink & R. Kay (2009), Agent Cognitive Capability and Orders of Emergence, In Trajkovski, Goran and Collins, Samuel (eds), "Agent-Based Societies: Social and Cultural Interactions", http://epubs.surrey.ac.uk/1606/1/fulltext.pdf
[Goldstein, 1999]	Jeffrey Goldstein (1999), Emergence as a Construct: History and Issues, Emergence: Complexity and Organization 1 (1): 49–72, http://www.wu.ac.at/am/Download/ae/Issue_1-1.pdf#page=50
[Heylighen, 1995]	F. Heylighen (1995): Downward Causation, in: F. Heylighen, C. Joslyn and V. Turchin (editors): Principia Cybernetica Web (Principia Cybernetica, Brussels), URL: http://pespmc1.vub.ac.be/DOWNCAUS.HTML

[Horgan, 1993]	T. Horgan, From supervenience to superdupervenience: meeting the demands of a material world, *Mind* 102, pp. 555-586, 1993. http://findarticles.com/p/articles/mi_m2346/is_n408_v102/ai_14640647/?tag=mantle_skin;content
[Hoyningen & Lohse, 2009]	P. Hoyningen-Huene and S. Lohse, Emergence: Postulates and Candidates, In: [2009] EPSA09: 2nd Conference of the European Philosophy of Science Association (Amsterdam, 21-24 October, 2009) - EPSA 2009 Contributed Papers. http://philsci-archive.pitt.edu/5209/1/hoyningen.lohse_emergence(final).pdf
[Johansson, 2002]	Ingvar Johansson, The asymmetries of property supervenience, In Sten Lindström and Pär Sundström, eds., Physicalism, Consciousness, and Modality, Umeå Preprints in Philosophy 2002:1, Department of Philosophy and Linguistics, Umeå University, pp. 95-124. http://hem.passagen.se/ijohansson/Superveniens1.htm
[Johnson, 2006]	C.W. Johnson (2006). What are emergent properties and how do they affect the engineering of complex systems? Reliability Engineering and System Safety 91:1475-1481, http://www.sciencedirect.com/science/article/pii/S0951832006000330
[Kim, 1984]	J. Kim, Concepts of Supervenience, Philosophy and Phenomenological Research 45 (1984), pp. 153-176; reprinted and quoted from Kim's Supervenience and Mind: Selected Philosophical Essays, Cambridge: Cambridge University Press, 1993.
[Kim, 1993]	J. Kim, Supervenience as a Philosophical Concept, Metaphilosophy 21 (1990), pp. 1-27; reprinted and quoted from Kim's Supervenience and Mind: Selected Philosophical Essays, Cambridge: Cambridge University Press, 1993.
[Kim, 1998]	J. Kim (1998). Mind in a Physical World. Cambridge, MA: MIT Press.
[Kim, 2006]	J. Kim (2006), Emergence: Core Ideas and Issues, Synthese 151 (3):547-559.
[Koestler, 1969]	Arthur Koestler (1969), A. Koestler & J. R. Smythies, ed., Beyond Reductionism: New Perspectives in the Life Sciences, London: Hutchinson
[Lewes, 1875]	G.H. Lewes (1875), Problems of Life and Mind (First Series), 2, London: Trübner, http://openlibrary.org/works/OL2799045W/Problems_of_life_and_mind
[Moyer, 2000]	Mark Moyer, Weak and Global Supervenience: Functional Bark and Metaphysical Bite? Draft Date: 6/2/00, http://www.uvm.edu/~mmmoyer/papers/sv_long.pdf

[McLaughlin & Bennett, 2010]	McLaughlin, Brian and Bennett, Karen, "Supervenience", The Stanford Encyclopedia of Philosophy (Summer 2010 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/sum2010/entries/supervenience/
[Peterson, 2006]	G.R. Peterson (2006). Species of Emergence. Zygon, 41(3), 22.
[Rasmussen et al, 1996]	S. Rasmussen, N.A. Baas, C.L. Barrett, M.W. Olesen, Dynamical Hierarchies – A Summary, Artificial Life & Robotics (AROB), Oita, Japan, February, 1996. http://www.osti.gov/bridge/product.biblio.jsp?osti_id=216259
[Rockwell, 2002]	T. Rockwell, A Defense of Emergent Downward Causation, online paper at http://users.sfo.com/~mcmf/causeweb.html (date unknown but before 2002)
[Sanz, 2004]	Ricardo Sanz, CO[7]: Converging trends in complex software-intensive control, In: Sixth Portuguese Conference on Automatic Control. Faro, Portugal, 2004. http://www.apca.pt/~apca_docs/CONTROLO2004/controlo2004/papers/pdf0001.pdf
[Scholte et al, 2009]	J.J. Scholte, H.A.P. Blom, J.C. van den Bos and R.B.H.J. Jansen, Management of ATM performance in operational concept development and validation: a case study, in Eighth USA/Europe Air Traffic Management R&D Seminar (ATM2009), http://www.atmseminar.org/seminarContent/seminar8/papers/p_102_APMM.pdf
[Shah et al, 2005]	Shah, A.P., Pritchett, A.R., Feigh, K.M., Kalaver, S.A., Jadhav, A., Corker, K.M., Holl, D.M., Bea, R.C., (2005.) Analyzing air traffic management systems using agent based modelling and simulation. In: 6th USA/Europe ATM R&D Seminar, Baltimore, USA. http://www.atmseminar.org/seminarContent/seminar6/papers/p_129_IAC.pdf
[SESAR, 2008]	SESAR Definition Phase, Task 4.2.1/D6 Sesar Safety Management Plan, Part of DLT-0710-421-01-00, version 1.0, May 2008.
[Sperry, 1964]	R.W. Sperry. (1964) Problems Outstanding in the Evolution of Brain Function, New York: American Museum of Natural History.
[Stephan, 2002]	A. Stephan, Emergentism, irreducibility, and downward causation, Grazer Philisophische Studien 65, pp. 77-93, 2002
[Stroeve et al., 2011]	S.H. Stroeve, H.A.P. Blom, G.J. Bakker, Contrasting Safety Assessments of a Runway Incursion Scenario by Event Sequence Analysis versus Multi-Agent Dynamic Risk Modelling, Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM 2011), June 2011

[Wyss, 2004]	Peter Wyss, Two aspects of emergence, Richmond Journal of Philosophy, Volume 8, Winter 2004, http://www.richmond-philosophy.net/rjp/back_issues/rjp8_wyss.pdf
[Zarboutis & Wright, 2006]	Nikos Zarboutis and Peter Wright, Using Complexity Theories to Reveal Emerged Patterns that Erode the Resilience of Complex Systems, 2nd Symposium on Resilience Engineering, Juan-les-Pins, France, November 2006, pp. 359-368, http://www.resilience-engineering.org/REPapers/Zarboutis_Wright_R.pdf