Creativity: Surprise and abductive reasoning

Maria Eunice Quilici Gonzalez
gonzalez@marilia.unesp.br
Pos-graduation in Cognitive Science and Philosophy of Mind,
Philosophy Department,
UNESP, Av. Hygino Muzzi Filho 737, Marília, SP, 17525-900, Brazil

Willem (Pim) Ferdinand Gerardus Haselager
w.haselager@nici.kun.nl
www.nici.kun.nl/~haselag
Artificial Intelligence/Cognitive Science, Nijmegen Institute for Cognition and Information,
University of Nijmegen, Montessorilaan 3, 6525 HR, Nijmegen, The Netherlands
Pos-graduation Cognitive Science and Philosophy of Mind, Philosophy Department,
UNESP, Av. Hygino Muzzi Filho 737, Marília, SP, 17525-900, Brazil

Abstract
This paper investigates creativity focusing on the nature of abductive reasoning, as originally formulated by Peirce, situating it in the context of the theory of self-organization. An ancient question will be addressed: is it appropriate to investigate creative processes from a mechanistic perspective or do they involve subjective elements which cannot - in principle - be investigated from a mechanistic view? This question will guide our investigation, which has as an initial hypothesis that creativity starts with surprise and involves a self-organizing process in which abductive reasoning occurs allowing the expansion of well-structured set of beliefs. This process is considered a part of the establishment of habits in self-organizing systems. We argue that a deeper understanding of how self-organizing processes involving abductive reasoning may take place in creative systems could elucidate the complex debate about the mechanical versus non-mechanical ingredients of creativity.

Keywords
abductive reasoning, conceptual space, creativity, emergence, habit formation, order parameters, self-organization, surprise

Introduction
It is our objective in this paper to investigate the nature of abductive reasoning, as originally formulated by Peirce (1931-1958) in the context of an old opposition between mechanistic versus anti-mechanistic approaches to creativity. We will start by addressing this opposition from the perspective of philosophy and cognitive science, focusing on Turing's (1950) and Boden's (1990; 1996) conceptions of creativity. We will present Peirce's characterization of abductive reasoning, relating it to creativity, surprise and the activity of habit formation. Finally, we suggest an analysis of creativity, based upon the theory of self-organization (Debrun, 1996; Haken, 1977, 1983) in order to explain how surprise can lead to the abandonment of inadequate habits and the formation of more adaptive ones.

Creativity and the mechanicism versus anti-mechanicism controversy

There is a tendency in our every day discussions, and even in philosophical debates about the nature of creativity, to reject any sort of mechanistic approach to novel ways of reasoning based upon the idea that creativity involves subjective or irrational elements. Popper (1961: 31), for example, says:

'The question of how it happens that a new idea occurs to a man - whether it is a musical theme, a dramatic conflict, or a scientific theory - may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge.... My view of the matter, for what it is worth, is that there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every discovery contains "an irrational element", or "a creative intuition", in Bergson's sense.'

From this perspective, creativity involves an element of subjectivity that makes it difficult, if not impossible, to be analysed from an objective perspective based upon pre-established rules or mechanical laws. In opposition to this view, as is going to be explained in the next section, philosophers like Peirce developed a theory of the logic of creative processes stressing a distinction between reasons to suggest a new hypothesis (as a possible solution to a certain problem) and the motivations that make an individual choose specific strategies to solve a problem. In the same vein, cognitive scientists claim that there is no obvious reason why a mechanistic approach to creativity should not be seriously investigated.

This debate about objective and/or mechanistic versus anti-mechanistic approaches to creativity is not new. In the 19th century, for example, Lady Lovelace registered her position against the hypothesis that computers (i.e. Babbage's analytical engine) could help us to understand creativity. She argued that, because computers can do only what their programs allow them to do, they do not constitute good instruments to understand creative thinking. Latter on, in the early stages of artificial intelligence, Turing (1950) responded to Lady Lovelace's argument explicitly by sketching a mechanistic approach to the mind, in general, and to creativity, in particular. He suggested that a machine could solve problems by operating with a set of rules on the basis of which it could move from one state to the next erasing and printing symbols. He insisted that even though computing machines have to be programmed to perform specific tasks, they often surprised him in his expectation about their performances either because he was distracted or because he forgot the calculations involved in specific tasks. Furthermore, a universal machine that can simulate the behaviour of other Turing machines could perform a variety of complex tasks, thereby increasing the changes of surprising its programmer.

Although Turing (who was certainly a very creative thinker) conceives the possibility of machines surprising himself, he does not consider the possibility that machines themselves could be surprised. If we consider the possibility that the experience of surprise may actually constitute an important element of creative thinking then Turing seems to miss the point of Lovelace's objection. In other words, if we accept the hypothesis that creativity involves an element of surprise, inherent to the experience of novelty, then Lady Lovelace's remarks seem to be most perceptive: the absence of a capacity to perceive novelty in the flow of information processed by computers (preventing them of experiencing surprises) may not make them good instruments to understand creative thinking. One hypothesis to be considered here is that creativity is directly related to the ability of being surprised.

Conceptual spaces and creativity

The ability to be surprised can be related to problem solving activities in which explorations of problem spaces lead to the expansion of belief domains. A successful expansion of beliefs is initiated by (and then eliminates) the experience of surprise. General heuristics have been described to guide search processes in problem solving activities; they include strategies for examining, comparing, altering and combining concepts, strings of symbols, and the heuristics themselves. However, critics insist that, despite their general appearance, heuristics are shaped for specific purposes and computers incorporating them will only do what their programs tell them to do without any possibility of experiencing surprises: no real creativity is involved in their activities, so the history goes.

In cognitive science, researchers like Newell & Simon (1972) and Boden (1990; 1996) attributed little or no importance to the experience of surprise in creative thoughts. According to Boden (1990: 30), for example, surprise and value, though important, are not the main elements to consider an idea as creative:

'To be creative, is not enough for an idea to be unusual – not even if it is valuable, too. Nor is it enough for it to be a mere novelty, something which has never happened before'.

She proposes to analyse creativity in terms of exploration and transformation of conceptual spaces. These are, as she stresses, multidimensional structures organized in accordance with principles that unify a domain of thinking. Such principles constitute the generative system that underlies a certain domain and defines its range of possibilities. Explorations in domains of thinking often lead to the expansion of, and, more rarely, useful transformations in the structure of the generative system, which provides the basis for novelties.

An important contribution to the study of creativity in the domain of Cognitive Science is given by Boden's analysis of how exploration in conceptual spaces may lead to relevant novelties in the domain of music, visual arts, literature and science. In science, for instance, she investigates Mendeleyev's creative process with the proposal of the periodic table in the 1860s: initially he classified chemical elements in rows and columns according to their similar observable properties and behaviour. In his processes of classification he left some gaps, predicting that in the future new appropriate elements could be discovered to fill them. Several years latter such elements were discovered whose properties satisfied his predictions and, moreover his table led to a more powerful classification of elements in terms of atomic number. This, in turn, explained the gaps originally left by Mendeleyev in his original classification.

Examples like the above illustrate situations in which mapping and explorations in conceptual spaces may lead to the expansion and generation of new ideas. Sometimes changes in conceptual spaces may lead to radical transformations, and not just expansions, in the constraints that define them. In this context, Boden (1996:81) describes the structural changes found in the development of post-Renaissance Western music based on the generative system known as tonal harmony:

'From its origins to the end of the nineteenth century, the harmonic dimensions of this space were continually tweaked to open up the possibilities ... implicit in it from the start. Finally, a major transformation generated the deeply unfamiliar (yet closely related) space of atonality.'

After investigating the steps of the process of undermining the notion of 'home key', which

led to atonal music, Boden stresses that the final, culminating, transformation realized by Schoenberg was the adoption of different constraints (such as the usage of every note in the chromatic scale) to structure his atonal music. Schoenberg's creativity consisted not just in the rejection of a constraint, but also in the generation of new ones, which allowed new forms of musical compositions.

In summary, the essential contribution given by Boden to the understanding of the origins of new ideas lies in making clear that conceptual exploration leading to the expansion or sometimes to the transformation of the generative structure of a domain can be seen as a form of creativity. Moreover, she suggests two senses in which creativity should be described. The first focuses on the psychological aspects, which characterize a creative (individual) mind in its uniqueness:

'A valuable idea is P-creative if the person in whose mind it arises could not have had it before; it does not matter how many times other people have already had the same idea.'(Boden, 1996:76).

In contrast, the second sense of creativity stresses its historical characteristics: 'A valuable idea is H-creative if it is P-creative and no one else, in all human history, has ever had it before' (Boden, 1996:76).

Acknowledging the role played by cultural values in the classification of ideas as creative or not (given that worthless new ideas are not considered creative), Boden is mainly concerned with the question of 'what does it mean to say that an idea could not have arisen before?' She distinguishes ideas that are merely novel ones - that can be produced by the same set of generative rules which produced other familiar ideas (as in Mendeleyev's classificatory procedures) - from radical and genuinely original ideas that cannot be produced in this familiar way (like Schoenberg's atonal music). The appearance of such new ideas presupposes going beyond the limitations of the pre-existing conceptual area in which they would not have found their natural space. As Boden (1996:78-79) stresses:

'the ascription of creativity always involves tacit or explicit reference to some specific generative system. It follows … that constraints - far from being opposed to creativity - make creativity possible. To throw away all constraints would be to destroy the capacity for creative thinking. Random processes alone, if they happen to produce anything interesting at all, can result only in first-time curiosities, not radical surprises.'

In this sense, creative thinking seems to presuppose: (a) some form of recognition of the principles and regularities that structure and, consequently, constrain a well-established conceptual space, and (b) a subsequent impulse to overcome them. As we are going to see in the next section, these two steps involved in creative thinking constitute the basis of abductive reasoning as proposed by Peirce. However, one relevant distinction between Boden and Peirce approaches to creativity is related to the role of surprise in the process of discovering new ideas.

Creativity, abductive reasoning and surprise

According to Peirce (1931-1958), the production of habits constitutes the main activity of the mind. A network of strong habits, in turn, gives place to beliefs on the basis of which novelties may be detected, accompanied by the experience of surprise:

'For belief, while it lasts, is a strong habit, and as such, forces the man to believe until some surprise breaks up the habit.' (CP 5.524).

On the basis of well-established beliefs, embodied minds operate in everyday, habitual, life with expectations that allow the anticipation of events. Fortunately, these expectations are fulfilled most of the time, but given the dynamical character of life sometimes there is a conflict between well-established beliefs and the environment in which organisms exist. This conflict produces in the mind a surprising effect, which according to Peirce, may be active or passive. The first occurs 'when what one perceives positively conflicts with expectations'. The second kind of surprise occurs 'when having no positive expectations but only the absence of any suspicion of anything out of the common, something quite unexpected occurs, - such as a total eclipse of the sun which one had not anticipated' (CP 8.315).
Under the effect of a surprise, which confronts expectations produced by well-established beliefs, several doubts appear in the mind stimulating it to inquiry until the experience of surprise disappears. Given the nature of beliefs (understood as strong habits), doubts will not disappear easily. They will persist until a new set of beliefs arises, transforming the surprising situation into 'a matter of course'. It is in this process of expansion or abandoning of well-established beliefs, triggered by the experience of surprise, that creative thinking may happen.

One of the greatest contributions of Peirce for the study of creativity resides in his analysis of the process of generation of new beliefs present in creative reasoning. In a well-known passage, he suggests the following logic description of creative reasoning, known as abductive reasoning:

A surprising fact, C, is observed.
But if H were true, C would be a matter of course.
Hence, there is reason to suspect that H is true. (CP 5.189)

Importantly, Peirce emphasizes that it is the confrontation with an unexpected phenomenon that sets off the process of abduction (CP 2.776) and that the aim of abduction is to avoid further surprises:
'What is good abduction? What should an explanatory hypothesis be to be worthy to rank as a hypothesis? Of course, it must explain the facts. But what other conditions ought it to fulfill to be good? The question of the goodness of anything is whether that thing fulfills its end. What, then, is the end of an explanatory hypothesis? Its end is, through subjection to the test of experiment, to lead to the avoidance of all surprise and to the establishment of a habit of positive expectation that shall not be disappointed.' (CP 5.197).

Peirce stresses that abductive inference, underlying creative thinking, does not provide absolute guarantees about its correctness: abduction is a fallible, but extremely useful form of reasoning guiding the mind in its attempt to free itself from doubts. He expresses his wonder about the organism's tendency to 'err' in the right direction when acting on the basis of this 'natural instinctive faculty':
'This Faculty is at the same time of general nature of instinct, resembling the instincts of animals in its so far surpassing the general powers of our reason and for its directing us as if we were in the possession of facts that are entirely beyond the reach of our sense. It resembles instinct too in its small liability to error: for though it goes wrong oftener than

right, yet the relative frequency with which is right is on the whole the most wonderful thing in our constitution.' (CP 5.173; see also Peirce's remarks on the 'affinity' between ideas and nature's ways in CP 2.776)

As is well known, Peirce does not restrict abductive reasoning to the human mind. He also insists that thought is not an exclusive capacity of the human brain:
'Thought is not necessarily connected with a brain. It appears in the work of bees, of crystals, and throughout the purely physical world; and one can no more deny that it is really there, than that the colours, the shapes, etc. of objects are really there.' (CP 4.551).

At the same time, his use of terms like 'instinct' (e.g. in CP 7.220: 'the existence of a natural instinct for truth is, after all, the sheet-anchor of science.') in relation to abductive reasoning could be taken as an indication that he would have considered abductive reasoning as a property of naturally evolving organisms, and not of artificially created, mechanical, systems. Moreover, as mentioned above, Peirce related abduction to the experience or 'feeling' of surprise:
'Abduction makes its start from the facts, without, at the outset, having any particular theory in view, though it is motivated by the feeling that a theory is needed to explain the surprising facts.' (CP 7.218).

One reason for the continuous recurrence of the 'mechanism versus anti-mechanism' debate on creativity is that the ability to experience or feel something is considered by many to be still outside the capacity of present day artificial mechanical systems, thus making it impossible for them to engage in genuine processes of abduction and creativity.
In the next section, we will investigate this topic from the perspective of the theory of self-organization and the general system theory. But before doing that, let us summarise our main points so far: According to Peirce the role played by abductive reasoning in creative thinking is directly related to the experience of surprise, which initiates the process of generation, change and expansion of beliefs understood as a form of strong, well established habits. This process, we claim, shares several similarities with the process of expansion of conceptual spaces suggested by Boden, as presented in the previous section. The main similarity between the ideas of Peirce and Boden in this respect resides in the supposition that the mind, in its tendencies to operate with well-established forms of beliefs, sometimes experiences anomalies or problems. This experience may initiate the abductive process of search of those possible hypotheses that, if true, could resolve the problem in question eliminating the feeling of surprise. However, Boden, in contrast with Peirce, does not seem to attribute any importance to the experience of surprise in the creative process. Would this difference make any difference (to use Bateson's notion of information) for the investigation of mechanical and organic creativity?

Self-organization and the changing of habits
We have selected the experience of surprise as a possible candidate to help us address the polemic 'mechanistic versus non-mechanistic' debate on creativity. As argued in the previous section, surprise triggers the abductive process and should disappear when abductive reasoning is completed, i.e. when a hypothesis is found that explains the event initially considered anomalous. Thus, it seems that, according to Peirce, not only the

experience of surprise (produced by the perception of events that do not conform to expectations generated by well-established beliefs) constitutes the first step of abductive reasoning underlying creative thinking, but also that its absence should be a characteristic of a good (completed) abduction. On the basis of these presuppositions we would like to investigate the following question: do our present day artificial mechanical systems, such as computer programs that model processes of scientific discovery, connectionist machines that can acquire and change habits or robots that interact with dynamic environments, have the ability of experiencing surprises?

However, a more basic, question needs to be answered before this polemic can be properly approached: if creativity involves, in general, a set of well structured beliefs, which operate as habits that constraint the flow of expectations, in relation to which surprising facts can be experienced, how are these constraints produced in the first place? In what follows we are going to examine this question from the perspective of the Theory of Self-Organization[1] (TSO).

As pointed out in Debrun (1996; see also Gonzalez and Haselager, 2002), the label self-organization' refers to a process through which new forms of organization emerge mainly from the dynamic interactions between elements of a system without any a priori plan or central controller. Two basic phases are generally described in self-organizing processes. The first, known as primary self-organization, involves the encounter between organic or inorganic elements. Initially separated (or with independent behaviours) they get together, ideally by chance, initiating an interaction amongst themselves in such a way that they become coordinated and interdependent. In this primary phase, spontaneous interactions may give place to new structures or distinct forms of organizations. In Ashby's (1962:266) words, this process of order formation is '...self-organizing in the sense that it changes from separated parts to parts joined' without the presence of any kind of pre-established program.

In cognitive science, examples of primary self-organization are found in processes of pattern formation in neural networks. These patterns emerge from the interaction of a large number of neuron-like units that send excitatory and inhibitory signals to one another. Even though the activity of individual units may be governed by local rules, the emergent overall pattern, produced by a collective effect, is not rule governed. In other words, even though each singular unit plays a role in the organizational dynamics of the net, the property that really matters to its final emergent organization of patterns is a *collective*, self-organizing, one.

Another important aspect of TSO is that the dynamic interaction amongst the system constituents may allow the emergence of an order parameter. As characterised by Haken (1999), order parameters (or 'collective variables') constitute high-level patterns that result from the interaction between low-level components in a system. Once created, they constrain and control, in a circular feedback way, the behaviour of the low-level components, which, in turn, may change the high-level order parameters, etc. The dynamics of stable order parameter formation characterises the second phase of self-organization.

As indicated by Gonzalez (2000), the dynamics of order parameters formation distinguishes artificial neural nets from traditional machines. Initially, the net's units have little or no relevant interdependent relations amongst themselves. With training, primary self-organization may take place; each unit is affected by other units, forming patterns of

connectivity. Their organizations emerge mainly from the dynamics of competition, co-operation and adjustment established between the net's units and informational patterns available in the environment. Temporarily, different forms of patterns may emerge from the primary process of self-organization, but only one of them will evolve. In order to do so, neural nets have to acquire the ability to allow the emergence of stable order parameters.
In those cases in which such a condition is fulfilled, and order parameters are formed, we may look at neural networks as interesting tools to help us to understand creativity. Their interesting aspect come from the possibility of considering the process of order parameter formation as a way of characterizing the process of habit formation, which, once established, will constrain the behaviour of the system in which it appears. Taking into consideration this hypothesis, we claim that the process of pattern generation can be seen, in the context of abductive systems, as a mechanism of habit formation that occurs in the primary phase. During the second phase of the self-organizing process these patterns become, acquiring the status of beliefs.

According to this perspective, abductive systems, immersed in a dynamic environment, will develop the ability of creating expectations and perform actions guided by them. In normal conditions, the majority of these actions are performed in accordance with environmental constraints that constitute an important element of the self-organizing process of order parameter formation. However, given the complexity and immense diversity of elements that characterises the dynamic environment-organisms couplings, these sometimes get out of phase. Disorder and instability characterise this stage in which old habits and order parameters may show inadequate to deal with novelties. Under these conditions, as suggested by Peirce, the mind will struggle in order to recover its habitual state of expectation fulfilments. This struggle may take several forms according to the mechanisms of adjustments available, which will allow a system to be 'in phase' again with its environment. Creative systems will look for new habits or hypothesis that will possibly explain or dissipate the experience of anomalies. Less creative systems, with poor mechanisms of adjustments, may struggle to death holding up to previously established habits, beliefs or explanatory hypotheses.
In summary: the central hypothesis of the present paper is that creativity is a self-organizing process in which abductive reasoning occurs allowing the expansion of well-structured beliefs. We assume that abductive systems operate, in general, with a set of stable beliefs, ordered in accordance with order parameters established by secondary self-organization. In this context, expectations are created against the background of which anomalies (i.e. surprising events) can be detected, which disturb and sometimes interrupt the flow of normal behaviour. This initiates a new primary phase of self-organization in which new habits are formed and dispositions cluster together. In the second phase, habits may be stabilized and new beliefs established as candidates to explain the detected anomalies. So, when stable beliefs (structured through secondary processes of self-organization) are disturbed, abduction takes place. This may start another instantiation of the secondary phase of self-organization leading to the creation of new habits. In case of more severe disturbances, a new phase of primary self-organization may be initiated, but the constraints in relation to which surprising facts can be experienced are produced especially by secondary self-organization. In cases of successful creative processes, new beliefs develop in response to disturbances or perturbations (surprising events) through the initiation of a new cycle of primary and secondary self-organization.

Interestingly, the theory of self-organization provides many examples of systems with a well-established order parameter value (i.e. a stable behavioural pattern that constitutes a habit) that can suddenly jump to quite a different state after encountering a disturbance. This sensitivity to perturbations is a central characteristic of dynamical systems that are on the verge of changing the order parameter. We suggest that these changes in order parameters provoked by disturbances can be interpreted as the breakdown of old, and the formation of new, habits.

Taking this perspective provides some interesting results. First of all, the question as to how a system can generate hypotheses that are ´frequently right´, as Peirce puts it, is no longer a big mystery. It is, ultimately, the collective interaction of the components of the system in response to aspects of the environment that results in the higher-level order parameter. Thus, a new order parameter (interpreted here as an abduced hypothesis or habit) may arise out of the history of the interactions of components within the system as well as out of the history of the system's interactions with the environment. From this perspective it is hard to see how the resulting order parameter could not be relevant, grounded as it is in the process of ongoing couplings between the system and its surroundings.

Secondly, the pattern produced by a system governed by the new order parameter can produce the aspect of novelty, often found in creativity, in the sense that it need not be related to the old pattern in a straightforward way. Even initially small changes in the order parameter can ultimately give rise to wildly different stable behavioural patterns. A small difference in the behaviour of, for instance, an organism (under the influence of perturbations) can lead to different feedback from the environment which leads to slightly different behaviour once again, leading to different feedback, etc. This circularly coupled process can set off a cascade of organism-environment interactions that lead to behavioural patterns that have no direct connection to the behaviours displayed before the disturbance. Responses to perturbations can lead to such radically different behavioural patterns that they transcend the confines of the traditional set of beliefs, thus necessitating a transformation of this set in order to encompass the new behaviour. Note, again, that there is no reason to wonder how the new behavioural pattern (despite its originality) can be adaptive, since both the organism and the environment have added their share to it, collectively sculpturing it, so to speak, into its final stable form.

Mechanic systems and the dimensions of surprise

So far we have argued that creative thinking can be understood as a self-organizing process in which abductive reasoning occurs allowing the expansion of structured beliefs. We submit that creativity as involving abductive reasoning can be understood from this dynamic perspective. Through secondary self-organization a system behaves within the confines of a stable set of order parameters (habits). Under certain circumstances the appearance of disturbances provides sufficient conditions for the system to change its behavioural mode into an entirely new and different stable pattern. Thus, abduction can be partially understood as a self-*re*organization of the system into a new order parameter in response to perturbations by the environment. We say 'partially' because, as mentioned, abduction involves, in general, the experiencing of surprise. In the case of humans, beliefs are socially, biologically and historically created and sustained, requiring a background of

an intricate web of interconnected habits. But, in the case of artificially created creatures, what would be the equivalent of this?

A first step to resolve this question is to note that causes for surprise can differ substantially in nature. Noticing a puddle of water in an otherwise completely dry street may constitute a surprise, but the nature of the surprising event seems far removed from e.g. a solar eclipse. Secondly, surprise comes in a number of experiential varieties. One may compare the mild annoyance a person feels when a colleague doesn't appear on time for a meeting to the total confusion that a completely unprepared person experiences when witnessing a solar eclipse. Obviously, both aspects, the nature of the surprising event (the type of disturbance) and the experiential effect (the impact, the experienced intensity of the surprise), are related and should be thought of as constituting different dimensions of surprise.

What we take to be great cause for caution in speaking about surprise in artificial systems (computer programs, robots) is that they appear to be extremely limited in both the type of disturbances they can notice and the experiential effects these disturbances can have on them. Let us consider a few examples; Thagard's (1988) PI, and Brook's (1991; 1999) reactive robots.

Thagard (1988; see also Holland et al. 1986) created a symbolic computational model PI (an acronym for Processes of Induction, but i.e. dealing with hypothetical induction or abduction). PI is an early and most valuable attempt to understand how abductive processes in scientific discovery could be modelled computationally. An interesting case investigated by Thagard and modelled by PI was the discovery of Vitruvius (a famous Roman architect from the second half of the first century BC) of the wave theory of sound. As Vitruvius tried to describe the acoustic principles underlying the design of Greek amphitheatres he noted (as undoubtedly many did before and after him) that sound propagates and reflects. He wondered how to explain this phenomenon, which initiated an abductive process, resulting in his explanatory hypothesis of sound travelling across distance like a wave. Vitruvius explicitly noted the similarities between the propagating characteristics of sounds and water waves (Holland, et al. 1986: 339-340) and PI simulated various ways of making this connection, resulting in the explanatory hypothesis that sounds travel like waves.
A consideration of PI (for more details, see Haselager, 1997: 100-120) shows that on both dimensions it is hard to speak of PI experiencing any kind of surprise. In relation to detecting the nature of the event, PI is totally reliant on pre-given, symbolic descriptions of the most important aspects of the situation. PI is not equipped with any sensory systems and does not face the problem of having to notice the relevant information in the first place (amidst all other kinds of information that might be available). PI does not have to note the event, for whatever is relevant about the event is spoon-fed into it. In relation to the second dimension, the intensity of the experience, one need not delve into the issues of qualia or consciousness in order to determine that PI does not experience any surprise. In fact PI could be anything but surprised, as it was totally prepared for the problem at hand by means of its knowledge structure. PI's total knowledge base (Thagard (1988: 29) mentions 60 rules or concepts as the upper limit for PI) was pre-constructed to deal with Vitruvius' problem and little else besides. One might say that PI was having all its habits being ready to deal with the specific problem it was constructed for, and it would have been 'surprised' (in the sense of exhibiting total apathy, not in the sense of abducing anything) only if its input contained any other information than

that for which it was prepared. In all, as Thagard himself indicates, PI does not by itself find problems to solve. There is no account of when PI should start abducing (Thagard, 1988: 175-176). From our perspective, this means that surprise plays no role in PI's functioning. That, in turn, makes claims about its abductive and creative capacities to a significant extent metaphorical.

More recent models are presented by a very influential approach in robotics, the so-called subsumption architecture or reactive paradigm of Brooks (1991; 1999; see also Murphy, 2000). In comparison to PI, reactive robots occupy an almost complete opposite place in the spectrum of AI. Instead of being immobile, knowledge-driven inference machines, reactive robots are operating without any explicit central knowledge consultation, inferential processes or model construction. Instead, they consist of behavioural layers that connect input directly with output, constituting a behavioural pattern or habit. The layers do not consult each other by means of representation exchange, but instead compete for dominance by means of inhibition or suppression mechanisms. The overall behaviour exhibited by the robots is an emergent result of the interactions between the layers, both among themselves and with the environment.
This approach to robotics has become very popular because the robots are capable of interacting directly with the world, solving problems by means of their hardwired behavioural capacities or habits. Here too, however, one can establish that surprise, in any serious sense of the word, is absent, without having to deal with issues regarding qualia or consciousness. Here the capacity to be surprised is lacking not because the robots are fully pre-tuned to the anomaly to be solved (as in the case of PI), but because they are, as it were, too 'empty' to be surprised. Because they are completely and continuously 'on-line' (that is, they are directly interacting with the environment, see Grush, 1997, Wheeler & Clark, 1999) nothing can surprise them. They have no capacity to 'step outside' of the situation in order to observe that something unusual may be blocking their habitual way of dealing with their environment. Whatever happens, the behavioural layers will continue with their struggle for dominance, and no recognition of a situation as problematic or anomalous is possible. As in the case of PI, then, but for totally different reasons, here too there is little space for claims about the abductive and creative capacities for these robots, because any possibility for surprise is lacking.

Conclusion

We have presented some ideas on the nature of creativity and the theory of self-organization. We submit that combining the notions of habit, surprise, abduction and order parameter may open new possibilities for research on creativity in the domain of cognitive science. From this perspective, creativity can be seen as intrinsically related to the continuous process of breaking up habits and acquiring new ones through abduction. The abductive process, in turn, can be understood as the formation of new order parameters under the influence of surprising disturbances. In this context, it seems that the contrast mentioned between creative and mechanistic processes can be formulated with greater precision. The original contrast between mechanic and non-mechanic systems is not completely illuminating because, like any cognitive system, creative minds too deal with processes that could be called mechanical, particularly in the production of habits that help with the dynamic interaction between environment and organism. However, creative systems not only have the capacity to establish habits but, in addition, have the capacity to

experience surprise and then are able to self-organize to dissolve, temporarily, this feeling of surprise, through the production of new habits. The common element to different creative systems seems to be the capacity of generating new habits. In living beings, the ability to create and change habits allows organisms to act in order to further their own survival and to adjust their behaviour in accordance with environmental requests, changing the environment and being affected by it in a circular feedback way.

In the beginning of this paper we raised the question whether machines themselves could be surprised (in contrast with machines surprising us). As we indicated surprise is important because it is the realization that something is blocking the usual path or that something unusual is happening that initiates abduction. Combining this with Boden's suggestion that constraints make creativity possible, we would like to suggest that the surprise comes from the sudden realization that the constraints exist. The surprising events (i.e. anomalies) reveal the existence of the constraints that normally and often invisibly guide our habitual actions and experiences. In this circumstance, creative systems adjust their behaviour by transcending the constraints or through abiding by them in unexpected ways, thereby drawing attention to the regularity itself. Once one becomes aware of the generative principles constituting the conceptual space, it is possible to transform them deliberately (and not just through random processes).

Ultimately, then, it seems that it is the capacity to experience surprise when habits are being thwarted that differentiates creative organisms from the purely mechanical systems that we considered in this paper. This capacity, in turn, seems to be directly related to the ability to adjust to the environment and to step out of problematic situations in order to realize that constraints exist. Investigations of the nature of adjustment in dynamic, self-organizing systems may open up new possibilities for understanding creativity, which at present still constitutes one of the most intriguing mysteries of evolving minds[2].

References

Ashby, W.R. (1962). Principles of the self-organizing system. In Principles of self-organization, H. Von Foerster. & G. W. Zopf, Jr. (eds.), 255-278. London: Pergamon.

Boden, M. (1990). The creative mind. London: Sphere Books.

__(1996). What is creativity? In Dimensions of creativity, M. Boden (ed.), 75-117. London: MIT Press.

Brooks, R. (1991). Intelligence without representation. Artificial Intelligence, 47, 139-159.

__(1999). Cambrian Intelligence. Cambridge, MA: MIT Press

Debrun, M.A. (1996). A idéia de auto-organização. In Auto-organização – estudos interdisciplinares, Coleção CLE. Vol. 18, M. Debrun, M. E. Q. Gonzales & O. Pessoa Jr. (eds.), 3-24. Campinas: CLE.

Gonzalez, M.E.Q. (2000). The self-organizing process of distributed information: a way out of the mind-body problem? In Proceedings of the 5th Brazilian – International Conference on Neural Networks, C.H.C. Ribeiro & F.M.G. França (eds.), Rio de Janeiro: CD-Rom.

Gonzalez, M.E.Q. & Haselager, W.F.G. (2002). Abductive reasoning, creativity and self-organization. Cognitio 3, 22-31.

Grush, R. (1997). The architecture of representation. Philosophical Psychology, 10 (1), 5-23.

Haken, H. (1977). Synergetics: An introduction. Berlin: Springer Verlag.

__(1983). Synergetics. Berlin: Springer Verlag.
__(1999). Synergetics and some applications to psychology. In Dynamics, synergetics, autonomous systems: Nonlinear systems approaches to cognitive psychology and cognitive science. W. Tschacher & J.-P. Dauwalder (eds.), 3-12. London: World Scientific.
Haselager, W.F.G. (1997). Cognitive science and folk psychology: The right frame of mind. London: Sage.
Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). Induction. Cambridge: MIT Press.
Knyazeva, H. & Haken, H. (1999). Synergetics of human creativity. In Dynamics, synergetics, autonomous systems: Nonlinear systems approaches to cognitive psychology and cognitive science. W. Tschacher & J.-P. Dauwalder (eds.), 64-79. London: World Scientific.
Murphy, R.R. (2000). Introduction to AI Robotics. Cambridge, MA: MIT Press.
Peirce, C. S. (1931-1958). The collected papers of Charles Sanders Peirce. C. Hartshorne, P. Weiss and A. Burks (eds.). Cambridge, MA: Harvard University Press.
Popper, K. (1961). The logic of scientific discovery. New York: Science Editions.
Newell, A. & Simon, H.A. (1972). Human problem solving. New Jersey: Prentice-Hall.
Thagard, P. (1988). Computational philosophy of science. Cambridge: MIT-Press.
Turing, A. (1950).Computing machinery and intelligence. Mind, LIX (236), 433-460.
Wheeler, M. & Clark, A. (1999). Genic Representation: Reconciling Content and Causal Complexity, British Journal for the Philosophy of Science, 50(1), 103-135.

---

[1] In a recent paper, Knyazeva & Haken (1999) analyse creativity from the perspective of self-organization and emphasize the importance of order parameters and the reorganization of problem (conceptual) spaces. However, they do not address the role of either abduction or surprise in this process.

[2] The authors would like to thank Lauro F.B. da Silveira, Candida, Mariana and Jonatas Manzolli for their many suggestions and support during the preparation of this paper. We would also like to thank FAPESP for funding the research activities of W.F.G. Haselager in Brasil and NICI for their permission for several prolonged stay at UNESP, Marilia, SP, Brazil. This paper is an improved extension of an earlier paper that was published in Portuguese in Cognitio.