# Superstition in the Machine

Alexander Riegler

Center Leo Apostel, Vrije Universiteit Brussel
Krijgskundestr. 33, B-1160 Brussels, Belgium
ariegler@vub.ac.be

**Abstract.** It seems characteristic for humans to detect structural patterns in the world to anticipate future states. Therefore, scientific and common sense cognition could be described as information processing which infers rule-like laws from patterns in data-sets. Since information processing is the domain of computers, artificial cognitive systems are generally designed as pattern discoverers.

This paper questions the validity of the information processing paradigm as an explanation for human cognition and a design principle for artificial cognitive systems. Firstly, it is known from the literature that people suffer from conditions such as information overload, superstition, and mental disorders. Secondly, cognitive limitations such as a small short-term memory, the set-effect, the illusion of explanatory depth, etc. raise doubts as to whether human information processing is able to cope with the enormous complexity of an infinitely rich (amorphous) world.

It is suggested that, under normal conditions, humans construct information rather than process it. The constructed information contains anticipations which need to be met. This can be hardly called information processing, since patterns from the "outside" are not used to produce action but rather to either justify anticipations or restructure the cognitive apparatus.

When it fails, cognition switches to pattern processing, which, given the amorphous nature of the experiential world, is a lost cause if these patterns and inferred rules do not lead to a (partial) reorganisation of internal structures such that constructed anticipations can be met again.

In this scenario, superstition and mental disorders are the result of a profound and/or random restructuring of already existing cognitive components (e.g., action sequences). This means that whenever a genuinely cognitive system is exposed to pattern processing it may start to behave superstitiously. The closer we get to autonomous self-motivated artificial cognitive systems, the bigger the danger becomes of superstitious information processing machines that "blow up" rather than behave usefully and effectively. Therefore, to avoid superstition in cognitive systems they should be designed as information constructing entities.

**Keywords:** Action-selection, anticipation, constructivism, decision-making, information-processing, pattern search, philosophy of science, schizophrenia, superstition.

## 1 Preliminary Remark

In his report, one of the reviewers wrote that the submitted version of this paper lead him "*initially* […] into the wrong direction of thinking" [my emphasis]. *Voilà.* This is

M.V. Butz et al. (Eds.): ABiALS 2006, LNAI 4520, pp. 57–72, 2007.

what this paper is about: picking up cues and running off in a direction that is determined by one's own experiential past. On a philosophical level the paper explores the relationship between rational thinking, anticipation, and superstition by building on the philosopher Kant's idea that "objects must conform to our knowledge" [32] rather than the other way around, which considers knowledge a mirror of the state of affairs. This paper is intended as a criticism of representationalist "third-person" modeling, of the attempt by humans to create intelligent artifacts in their own image.

## 2 Introduction

While early AI was mainly concerned with symbolic computation that assumed a readily structured propositional environment, more recent streams emphasize the embodied dynamical nature of cognition, e.g., [56]. In this paper I address one of the main consequences of the embodiment paradigm, i.e., the question of the relationship between the cognitive agent and its environment and the potential danger of the view that the agent is informed by the environment, i.e., that the agent processes input information in order to generate output, or as Ulric Neisser [48] puts it, that cognitive subjects are "dynamic information processing machines." Since in the context of human beings it can be shown that this leads to superstition and mental disorders, it seems reasonable to prevent machines from this destiny by carefully crafting alternative design principles.

This paper starts with the extreme case of a structureless (amorphous) world. This shifts the focus of attention from structures "out there" (entities, events, etc.) to what goes on *inside* a cognitive being. Research in adaptive behavior and cognitive systems is, after all, interested in creating cognitive artifacts rather than artificial worlds. I proceed with arguing that, based on experimental findings, there is a close relationship between pattern discovery and superstition since humans and animals alike excel at finding structures where there are none. How can this be explained?

At first sight it seems that the ability to find structures and compress them into rules is rather useful for anticipating future states. This ability is usually called "inductive reasoning." However, in many cases, instead of anticipating states that become actualized in the future, the cognitive systems merely exhibits wishful thinking, also referred to as "superstition." The point is that both induction and superstition are carried out by the same cognitive apparatus: from its point of view ("first-person perspective") there is no difference. But why is our information processing not always successful? We can identify two classes of reasons: (L1) our cognitive equipment is *very* limited by a small short-term memory, conservative bias in problem solving ("set-effect"), and the illusion of explanatory depth. (L2) The combinatorial explosion of different ways to account for the links between entities and events is such that the computational effort required to compute them becomes intractable or NP-complete in the sense that the time required to solve them grows exponentially with the number of components.[1] So how can complex situations be *processed* by limited cognition?

[1] This was one of the main reasons why attempts of AI failed to scale blockworld scenarios up to real world situations.

This suggests that the cognitive being does not map structures from the "outside world" onto its cognitive equipment but rather creates structures in the first place. The construction process may be triggered by sensorial, proprioceptive or other, "internal" cues. A reader, for example, would not be able to read a novel without the faculty of creating lively, rich mental pictures out of a few letters on a page. However these pictures are composed of parts that already exist and which were constructed previously. The parts are put together in a particular way, creating anticipatory "checkpoints" between them. The function of the checkpoints is to verify the viability of the constructed chain. By constructing this chain, cognition is canalized into a particular direction making it possible to effectively control the combinatorial explosion in the sense of L2 (i.e., to prune the vast search space – a very well-known problem in artificial intelligence and cognitive science). In this sense, cognitive structures are projected onto the "external" world.

Superstition occurs when the cognitive system is exposed to L2, i.e., when it leaves the construction mode and tries to find sense in the flood of incoming data (in humans this search for new rules may be accompanied by a feeling of anxiety). However the chances are that the newly constructed rule is nothing but a bad guess.

The remaining part of the paper is concerned with providing empirical and argumentative support for the thesis that cognition is about information construction rather than information processing. I start by presenting arguments that make it clear that the structure of the "world" must be almost infinitely rich so that we can speak of an amorphous world. Then I cite empirical results from the psychological literature that suggest a close link between pattern detection, anticipation and superstition. Furthermore, I discuss decision making from the perspective of both information processing and information constructing. I conclude that by following self-constructed information we do not fall prey to arbitrariness or insanity. Consequently, I suggest applying these insights to the design of genuinely autonomous artificial cognitive systems in order to prevent them from "blowing up."

## 3 The Search for Patterns

A generally accepted working hypothesis in artificial intelligence and cognitive science is that common-sense thinking and science both attempt to infer rule-like laws from the patterns in data sets, i.e., to perform some sort of data-mining.[2] In philosophy of science we find this claim expressed, for example, in Ernst Mach's *economy of thought* [41] which states that the goal of science is "the simplest and most economical abstract expression of facts" [42]. Similarly, Herbert Simon defined (scientific) discovery as "detecting the pattern information contained in the data, and using this information to recode the data in more parsimonious form" [60]. He argued that computer programs can discover the recursive rules generating sequences of letters. However, as the following examples demonstrate (from [28]), even in the case of seemingly simple sequences it can be hard to find the respective rule that specifies the criterion for separating the sets of letters in the following three sequences.

---

[2] Science may do so with a higher degree of systematicity, with the goal of uncovering underlying causes.

(A, E, F, H, I, K, L, M, N, T, V, W, X, Y, Z) versus (B, C, D, G, J, O, P, Q, R, S, U)
(A, B, D, O, P, Q, R) versus (C, E, F, G, H, I, J, K, L, M, N, S, T, U, V, W, X, Y, Z)
(A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, W, X, Y, Z) versus (V)

In order to establish rules in the first example you need to focus on whether all the segments of the respective letter are straight. In the second example the distinction is based on whether the respective letter has at least one enclosed area. That the criteria don't need to be based on topological characteristics is demonstrated in the third example in which the letters are separated based on whether they are part of the Polish alphabet, which does not include the letter V. With a little bit of imagination even more unusual letter sequences can be formulated. How could a heuristically working AI program possibly consider all possibilities to find the appropriate rule?

In his influential book *The logical structure of the world,* Rudolf Carnap presents the thought experiment of two geographers – a realist and an idealist – who travel to Africa to discover a certain mountain. After they have collected empirical data about its location, height and other physical characteristics upon which they agree, they find themselves in disagreement with regard to how to interpret the data. For the realist, the mountain "not only has the ascertained geographical properties, but is, in addition, also real" while the idealist claims that "the mountain itself is not real, only our perceptions and conscious processes are real" [10]. Carnap's original intention was to show that epistemological theses do not have any value if they go beyond experience in the sense that neither proponent can give an "indication of the design of an experiment through which his thesis could be supported" [10]. Historically, Carnap's arguments were meant to support logical positivism. However, the thought experiment also demonstrates the apparent arbitrariness of how to read sense into experiential data, which, in contrast to general belief, makes the scientist rather than "nature" responsible for decisions. The conclusions are twofold. For philosophy of science, theories are plausible at best; and for artificial intelligence, the structure of the environment seems to play only a marginal role.

## 4 The Amorphousness of the World

Let us investigate the role of the environment a little further. As early as 1906, Pierre Duhem [18] claimed that observational evidence can never conclusively disprove a theory (or thesis) as any seemingly disconfirming observational evidence can always be accommodated to it (the so-called "underdeterminism" of theories, in modified form later known as "Quine-Duhem thesis"). As a result, there will be many competing theories trying to explain a given set of experimental data.

Philosophy of science is full of examples that support the underdeterminism theorem, e.g., [35]. However, one can easily form an idea of how vast the range of possibilities is by considering the abstract concept of a black box, i.e., an entity whose inner mechanisms are unknown to the outside observer. It is already extremely difficult to make inferences for a black box with four input, four internal, and four output states, all of which can be wired in any way [21]. The total number of possible configurations is $4\text{^}(4^4) = 2^{32}$, i.e., about $4 \times 10^9$. In other words, starting from an observational protocol one needs to test 4 billion different models to find the one that reproduces a recorded behavior. Usually, empirical data contains much more extensive protocols.

Going one step further, James McAllister points out that there is an arbitrarily large number of ways to explain data points: “Any given data set can be interpreted as the sum of any conceivable pattern and a certain noise level. In other words, there are infinitely many descriptions of any data set as ‘Pattern *A* + noise at *m* percent’, ‘Pattern *B* + noise at *n* percent’, and so on”[3] [45].

> Asked in what the phenomenon of planetary orbits consists, Kepler would have replied “In the fact that, with such-and-such a noise level, they are ellipses,” while Newton would have replied “In the fact that, with such-and-such a (lower) noise level, they are particular curves that differ from ellipses, because of the gravitational pull of other bodies.” Thus, phenomena – understood as the patterns in data sets that investigators choose to model – vary from one investigator to another [45].

McAllister [46] continues to argue that since any given data set can be interpreted as the sum of any conceivable pattern and a certain noise level, all the rules and patterns that a data set displays have equal status and can, therefore, be said to equally correspond to structures in the “world.” But if that world contains any structure, then it contains all possible structures, which is equivalent to exhibiting no structure at all, i.e., to being amorphous.

So in order to arrive at a decision, we let our choice be guided by pragmatic concerns such as simplicity, economy and elegance [55], or as Heinz von Foerster [22] put it,

> Only the questions which are principally undecidable, we can decide. Why? Simply because the decidable questions are already decided by the choice of the framework in which we are asked, and by the choice of rules of how to connect what we call ‘the question’ with what we may take for an ‘answer’. ... [We] are under no compulsion, not even under that of logic, when we decide upon in principle undecidable questions.

This insight seems to be in sharp contrast to the self-understanding of the natural sciences, which aim at quantitative exactness through systematicity [29]. For example, in classical physics quantitative empirical data enters into a model in such a way that the mutual influence among single data items can be determined computationally. This principle assumes a homomorphism between the structure of the phenomenon and the structure of the model, i.e., that the phenomenon can be reduced to an isomorphic copy of the model through applying a many-to-one transformation [5]. However, for a given phenomenon an innumerable amount of homomorphic models can be found through many-to-one transformations. All these models will exhibit the same behavior. The reverse inference from a given model (i.e., the scientific image) to the “true” structures of the phenomenon (i.e., reality) is therefore impossible. Rather, it is the human (scientist or other) who decides which structural pattern to read in the data in order to explain the phenomena of interest in terms of a scientific law. In this way, quantitative completeness is replaced by *qualitative* schemata [3]. He understands the

---

[3] The term “noise” refers to the mathematical discrepancy between a particular pattern and a given data set.

behavior of the homomorphic model of the "real" phenomenon rather than the phenomenon itself. That is, he is able to cope with a "suitable" simplification of the system's states (which is his invention) in order to make predictions "in the head." As a result, people often overestimate their knowledge of facts and procedure, a phenomenon that Frank Keil [33] calls the "illusion of explanatory depth," so that despite the vast incompleteness of their knowledge, most people think that they know far more than they actually do.

## 5 Superstitious Anticipations

Given the amorphousness of the world one is tempted to caricature human attempts to systematically generate knowledge as nothing more than having one's fortune told from the coffee cup; any arbitrary and unconstrained interpretation seems possible. Still, as we will see in the following examples from the psychological literature, it would be wrong to say that superstitions are "just" erroneous modes of information construction. Rather, superstition is the active attempt by the subject to coerce order into structurelessness.

As is known from the psychological literature, the search for patterns in one's stream of experience leads to various conditions. Klaus Conrad [13] coined the term *apophenia* [*Apophänie*] to refer to the experience of "unmotivated seeing connections" in random or meaningless data. Similarly, *pareidolia* is the erroneous or fanciful perception of a pattern or meaning in something that is actually ambiguous or random. Examples are the Rohrschach test and the alleged face on Mars based on a photo made from the surface of the planet. In general, humans are prone to the so-called *clustering illusion*, i.e., the tendency to associate some meaning to certain types of patterns that must inevitably appear in any large enough data set. According to Scott Huettel, Peter Mack and Gregory McCarthy, even if patterns are generated randomly their recognition is "obligatory, in that it occurs without any conscious attentional effort" [30]. In their experiments they confronted test subjects with a random sequence of squares and circles. The subjects were asked to press a button in their right hand when they perceived a square, and a button in their left hand for a circle. Occasionally, brief periods of seemingly non-random patterns appeared, such as a series of alternating circles or squares. Even though the subjects were instructed that they were seeing random sequences, their unconscious reacted when such a series was violated. This was demonstrated by functional magnetic resonance imaging (fMRI) scans of their prefrontal cortex, which revealed the changing activity in a distributed set of regions that are highly sensitive to the presence of and deviations from patterns. It seems that the subjects' motor behavior was primed, based on the belief of having discovered a pattern that would continue. In other words, humans construct a belief in a genuine regularity where there is none: "the recognition of patterns is an obligatory, dynamic process that includes the extraction of local structure from even random sequences" [30].

Such *compulsive* pattern-perception is not limited to event patterns as was shown in the experiments of Frédéric Gosselin and Philippe Schyns [26]. The authors stimulated the visual system of test subjects with unstructured white noise[4] that superimposes on the contours of a face. In 20,000 trials the subjects were asked to determine whether the face was smiling, which according to the instruction was the case in 50% of the presentations, even though in none of the presentations whatsoever did the face have a mouth. Still, in many cases the subjects were certain that the face was indeed smiling. Clearly, the anticipated pattern was projected onto (partially correlated with) the perception of the white noise.[5]

A possible explanation for the built-in tendency to perceive patterns is that it enables the subject to better react to a sequence of cues that signal a potential threat or a source of food. Therefore, the obsessive search for patterns not only serves as a basis for human superstitious behavior, it can also be found in the animal kingdom. B. F. Skinner's article on *Superstition in the pigeon* [61] is a classical description of how birds react in situations beyond their cognitive capabilities and therefore beyond their control. Skinner presented food at regular intervals to hungry pigeons with no reference whatsoever to their current behavior. Soon the birds started to display certain rituals between the reinforcements, such as turning two or three times about the cage, bobbing their head, and incomplete pecking movements. As Skinner noticed, the birds happened to be executing some response when the food first appeared and they tended to repeat this response if the feeding interval was short enough. In some sense, pigeons associated their action with receiving food and started to believe that it caused the food to appear.

In the early 20$^{th}$ century Bronislaw Malinowski noticed that islanders in the Pacific who fished offshore beyond the coral reef displayed many superstitious rituals and ceremonies to invoke magical powers for safety and protection, while inshore fishermen carried out their job with a high degree of rational expertise and craftsmanship [43]. It reflects the desire of humans to find causal explanations and to organize their experience in a meaningful manner [25] in order to make predictions based upon them. Ellen Langer [36] accounted for the tendency to apply superstition as a response to uncertainty by introducing the notion of "illusion of control," i.e., the belief that one can control or at least influence outcomes in situations under which one has no control: "If there is a universal truth about superstition, it is that superstitious behavior emerges as a response to uncertainty – to circumstances that are inherently random and uncontrollable" [65].

Engaging in superstitious behaviors such as displaying patterns of stereotyped behavior is closely linked to an illusion of control since the people engaging in these patterns may actually believe that *they* are controlling an outcome [57]. A prominent example of this perspective is *feng shui*, the ancient Chinese superstitious practice of placement and arrangement of space, which is claimed to achieve harmony with the environment. Even today, Chinese managers resort to this superstitious practice when they have to make important decisions as many of them find it difficult to cope with the unknown [63]. It helps them to reduce uncertainty-induced anxiety. In this sense,

---

[4] White noise is a static bit pattern that has equal energy at all spatial frequencies and does not correlate across trials.

[5] The Italian painter Leonardo da Vinci had already recommended looking at blotches on walls as a means of initiating artistic ideas: "If you look at walls covered with many stains … with the idea of imagining some scene, you will see in it a similarity to landscapes adorned with mountains, rivers, rocks, tree, plains, broad valleys, and hills of all kinds."

superstition provides an additional source of information which fills the void of the unknown or helps to deal with "information overload" by disentangling conflicting information to tip the scale: "superstition breaks the deadlock by indicating a superior alternative" [63]. One-third of Chinese managers rely so much on superstitious practices that they even neglect rational solutions. For others it is a stress-management tool, while for still others it has become an obsessive habit without which they will not feel at ease.

In the psychological literature, further links between superstitious thoughts and behaviors and obsessive-compulsive disorder (OCD) can be found, such as obsessive checking [23] and anxiety disorders in female test subjects [66]. Furthermore, as Peter Brugger [8] points out, the "ability to associate, and especially the tendency to prefer 'remote' over 'close' associations, is at the heart of creative, paranormal and delusional thinking." Again, a correlation between OCD and magical thinking, i.e., maintaining beliefs that defy culturally accepted laws of causality in general and the belief that certain thoughts or behaviors exert a causal influence over outcomes in particular, can be found [20].

The close connection between schizophrenia and creativity, i.e., the ability to think in terms of connections and links between entities belonging to different categories gives rise to the assumption that creativity has also close ties to superstition, which can be defined as the confusion of categories and core knowledge [37], i.e., building blocks that emerge early in human ontogeny and phylogeny [62]. This includes, among others, confusion of symbolic representations and the material objects they represent, the attribution of physical or animate entities to mental content, the ascription of independent existence to good and bad minds, which behave as animate entities by moving and initiating actions without external force, and thinking that badly placed furniture leads to crime and divorce.

The conclusion from these insights is that cognition, being the faculty to make decisions, depends on our ability to anticipate future events and states. This thesis, largely upheld and developed in cognitive science and AI, is not disputed. However, the origin of the anticipations is called into question: are they a result of information processing or alternatively do they emerge from information constructing? This question is legitimated by the fact that due to the amorphousness of the world any arbitrarily large number of underlying rules or laws for predicting events can be established. This leaves a cognitive system with uncertainty and even anxiety. In order to regain the feeling of control, cognition uses superstitious behaviors, which relieve the cognitive agent from the need to process the information overload from its environment. Therefore, for want of structure, superstition emerges as a response to uncertainty and may even transform into psychological disorders.

In the following section the consequences for decision-making and cognition are addressed.

## 6 Decision-Making and the Construction of Information

It has been argued that complex human thought and behavior are possible because cognition is able to simulate long sequences of responses and sensory consequences [27]. In a sense, cognition then consists of walking through a chain of decisions.

According to the traditional literature on decision-making, it is taken for granted that human cognition makes decisions based on careful considerations of the situational context. In economics, this idea has been condensed into "rational choice theory," which regards rationality as the only criterion a decision maker has to strive for. As pointed out above, there are reasons to assume that this is a misleading characterization, which is based on the assumption that human cognition basically functions as an information-processing device that extracts rules from incoming experience. This theory not only fails to take the limitations of (human and any other finite) cognition into account, it is also unable to deal with the amorphousness of the world. In artificial life research, this has been called the "what to do next" or action-selection problem [64, 31], i.e., formulating a mechanism that allows choosing an action in pursuit of a single coherent goal or several conflicting and heterogeneous goals.

Gerd Gigerenzer [24] argued that decisions must necessarily be adaptive, fast, and frugal if they are to ensure survival. In artificial intelligence, a similar paradigm has emerged focusing on behavior-based robotics. It does not rely upon *a priori* mathematical analysis of a given situation but rather on a *hic-et-nunc* strategy that takes *system-internal* drives into account rather than a sophisticated representation defined in terms of the programmer's semantic world, i.e., from the third-person perspective.

In order to understand how human and, consequently, artificial cognitive systems with limited processing capabilities can cognitively cope with an amorphous world, let us explore two examples from the animal kingdom which both question the notion of information-processing.

(1) Consider the behavior of an incubating goose that decides to use its bill to roll back the egg that has fallen out of its nest. Interestingly, it will continue its rolling behavior even if an ethologist takes away the egg [38]. It seems that the animal does not constantly screen its environment and filter out environmental changes. Rather, the environmental state becomes only important at certain, apparently evolutionarily important, *checkpoints* (which do not include the existence of ethologists). These checkpoints act as anticipations that determine whether an action that has been already started is on track with regard to a certain goal. As the psychological literature documents, human problem-solving, too, is dominated by a similar sort of conservative inflexibility that makes subjects repeatedly choose a once successful strategy irrespective of whether another, simpler, strategy might be better suited for new problems [19, 40]. Furthermore, human perception is determined by internal cognitive dynamics that only occasionally seek to verify certain anticipations about future input states, as shown in the sequential order of tactile object recognition [58]. Finally, it has been argued that the inability to ignore stimuli (i.e., low latent inhibition) can lead to mental illnesses such as schizophrenia [39].

(2) Consider a fly crawling over a painting of Rembrandt [44, 56]. It in no way *processes* the visual information presented in the painting, as from its perspective there is no painting whatsoever. Only the human observer may wonder which information *filters* the fly applies in order to *ignore* the rich informational input. The fundamental difference arises from the fact that human scientists and engineers from their third-person perspective (and lacking the first-order perspective of the observed systems, e.g., the fly) necessarily concentrate on the perceivable *output* of systems (such as the crawling of the fly). Cognitive systems, however, take actions in order to control and change their perceptive and proprioceptive *input*, e.g., they avoid the

perception of an obstacle or they drink to quench their thirst: “From the organism’s point of view only actions which feed back to the organism’s sensors can be observed […] Any other action which simply disappears in the environment cannot be observed by the organism” [53].

In the latter sense a cognitive system needs to be defined as an information *creator* rather than as an information *processor*. The latter is defined in terms of an input–output relationship: a given input or perception yields (directly or after some “calculation”) an action, i.e., the output is a function of the information in the input. Information-creating systems, however, control their input (perception) rather than their output (behavior), or as Powers [54] put it, “behavior is the process by which organisms control their input sensory data.” This can only be achieved by first creating information (e.g., chains of cognitive elements connected via anticipatory checkpoints) internally and then allowing those checkpoints to be (occasionally) matched against the input.

Historically this idea can be traced back to the cybernetic concept of homeostasis [9]. It says that a living organism has to keep its intrinsic variables within certain limits in order to survive. These “essential variables” (which include body temperature, levels of water, minerals and glucose, and similar physiological parameters, as well as other proprioceptively or consciously accessible aspects in higher animals and human beings [4]) represent the purpose of a system. In order to account for a wider range of systems, we have to make the following (rough) distinctions here:

(a) *Man-made vs. genuinely autonomous systems.* In man-made systems goals are defined by the human designer, whereas in genuinely autonomous systems the goals are constructed by the system itself. They are defined before the system reads its inputs and tries to accommodate for any deviations, or as William Clancey [11] expressed it, “what constitutes information for an organism cannot be given by a teacher, but must arise from the organism’s own organizing processes in interaction with its environment.” Since the agent controls its inputs the output can become quite unpredictable for an *external* observer [53]. It is therefore useful to retain the distinction between a first-person perspective (the one of the input-controlling system) and the third-person perspective of the external observer who can only make guesses about the self-defined goals of the system. Susan Oyama [51] maintains that there is no preformed or a priori information, but rather that every system constitutes its own information: information is bound “inextricably to a point of view,” i.e., the first-person perspective.

(b) *Simple vs. cognitive systems.* In simple systems (such as thermostats) the state of homeostasis is reached by the simple process of negative feedback. For example, in a thermostat, a given parameter, such as the temperature, is kept under control by appropriate counteractions, i.e., by turning the heating on or off depending on a certain reference value. Cognitive systems need to execute a certain sequence of actions in order to control and change their input state. This is a result of the fact that “perception and action arise together, dialectically forming each other” [12]. This renders the whole concept of representation doubtful as “we can walk through a room without referring to an internal map of where things are located, by directly coordinating our behaviors through space and time in ways we have composed and sequenced them before” [12]. Consequently sensory, cognitive, and motor functions can no longer be considered independent and sequentially working parts of the cognitive apparatus. And the latter can no longer be described as a computational device, the "task" of which is to acquire propositional knowledge about the mind-independent reality by processing information that is picked up from that reality.

Putting (a) and (b) together, *genuinely autonomous cognitive systems* can be characterized as systems that try to achieve self-determined goals. These goals are defined in terms of anticipatory checkpoints that glue together action sequences. In order to act as goals these action sequences must be constructed in the first place: the information they contain is constructed rather than representing processed input. The cognitive system then executes an action sequence as long as the checkpoints can be successfully met by the input. Any failure to do so causes the system either to find alternative chains of actions that more appropriately answer to the requirements of the checkpoints ("re-orientation") or to construct new chains that make more or less use of already existing elements. The lower the degree of reusability, the higher the degree of "perplexity" of the system. The vast number of different ways to form chains from components excludes random arrangements from being an option.

It is already clear in the case of simpler systems that action sequences have to be assembled in a hierarchical manner. While such a simple feedback loop may suffice for primitive intrinsic variables, higher order goals are accomplished in a hierarchical assemble of feedback loops in which each level provides the reference value for the next lower level: "The entire hierarchy is organized around a single concept: control by means of adjusting reference-signals for lower-order systems," as Powers [54] pointed out. So at higher levels the system controls the output of lower levels, at the bottom, however, it controls its perceptual input.

In cognitive systems action sequences must be arranged in a similar manner. Formally, the cognitive apparatus $P$ may consist of schemata $R$ that work over mental states $S$, $P = \langle R, S \rangle$.[6] Each schema $r \in R$ is a chain of checkpoints $c$ and actions $a$, $r = \{c \mid a\}^+$. We can extend this definition by allowing clusters of checkpoints and actions, respectively, $C = \{c\}^+$ and $A = \{a\}^+$, in order to account for multimodal perceptual entities and action sequences (e.g., the egg-rolling sequence in the example above), respectively. If, in addition, we allow recursions of the form $C = \{c \mid C\}^+$ and $A = \{a \mid A\}^+$ it follows that checkpoints and actions form *nested hierarchies* in which encapsulations are reused as building blocks on a higher level [1]. Elements can be functionally coupled as they make use of encapsulations and/or are part of encapsulations themselves. The encapsulation guarantees that, when new chains are being formed, partial solutions do not get lost, preventing the apparatus from making random guesses. By using a hierarchical arrangement the cognitive agent gains two advantageous aspects: speed and goals. According to Simon [59] "hierarchic systems will evolve far more quickly than non-hierarchic systems of comparable size" since the "time required for the evolution of a complex form from simple elements depends critically on the number and distribution of potential intermediate stable forms." In addition, the hierarchical composition of chains introduces a direction, a *goal*, "by the stability of the complex forms, once these come into existence." [59]

---

[6] Schemata are a popular explanatory vehicle for cognitive systems and have been widely discussed in the literature, e.g., Neisser [49], Drescher [17], Arbib [2], Bickhard [6].

In other words, goal-directed behavior can be explained as the execution of hierarchically nested chains composed of action sequences and checkpoints that glue them together. The checkpoint character of cognition defines decision-making and other cognitive acts as being based upon internal states rather than external states of affairs. For an observer, from her (third-person) perspective cognitive systems seem to make the proper decisions in response to the challenges of the environment. But from the (first-person) perspective of the organism, decisions are only the consequence of the internal cognitive dynamics. Since the observer is a cognitive system with her own internally defined goals, the respective goals of observer and observed do not necessarily coincide. This renders representational modeling difficult: how can the goal of the modeled cognitive system be defined in terms of modeler's goals? A design process like this requires a "God's Eye" perspective and is therefore, unless they trust in random guesses, out of reach for human scientists.

Reusing stable clusters of components, however, also has the disadvantage that the "right thing" may be done at the apparently "wrong moment." If we consider the superstitious behavior of pigeons or the rituals of the offshore fishermen, it is clear that exactly this characterizes superstitious behavior. None of the involved behaviors are new. Only the perceptual condition that triggers the execution of a certain sequence is different from the one that was originally associated with that behavior. In order words, superstition emerges when processed information is combined with already-constructed information (cf. the above discussion about the close link between superstition and creativity).

## 7 Conclusion

The arguments presented in this paper give rise to the assumption that artificial cognitive systems that are designed as information-processing devices are exposed to the danger of turning to superstitious behavior as their autonomy and cognitive competence increases. The arguments are based on the fact that in environments whose complexity transcends that of block- and microworlds[7] the number of possible rule-based explanations are NP-complete; any attempt to calculate them with the (very) finite means of the cognitive apparatus would render decision-making impossible. Making random guesses, however, leads to superstition and mental disorders.

Since Daniel Dennett's description of the frame problem [14], we know that robots that try to calculate every possible cause in their environment are doomed to remain inactive. In a sense, scientists are in no better of a situation. Johannes Kepler spent many years trying to find the regularity behind planetary movement by compressing the huge amount of position data collected by him and his teacher, Tycho Brahe [34]. Therefore it should not amaze us that "a scientific law or theory provides an algorithmic compression not of a data set in its entirety … but only of a regularity that constitutes a component of the data set and that the scientist picks out in the data." [47] In other words, some of these patterns "are taken by investigators as corresponding to phenomena, not because they have intrinsic properties that other patterns lack, but because they play a particular role in the investigators' thinking or theorizing."

[7] The environments targeted by many action-selection solutions, e.g., [64,31].

[45] These are the checkpoints in cognitive chains. Instead of processing the structures of the world we generate this structure by applying our cognitive operators [16], an insight which was originally proposed by Immanuel Kant's Copernican Turn [32].

When checkpoints are not met and the cognitive agent is forced to switch to information processing in order to construct new cognitive chains, superstition and mental disorders may emerge. The bigger the information overload the longer it takes to get back to the normal mode of cognition (cf. the amount of time necessary to find one's way through new information) and the higher the chance of constructing superstitious information.

To sum up, by taking an input-oriented perspective, we no longer need to assume that information is being processed but, rather, that it is constructed. That is, cognition is safeguarded from cognitive overload, which could otherwise lead to the illusion of control and, consequently, the creation of superstitious routines. Such an information-constructing paradigm emphasizes the primacy of the internal cognitive dynamics over influences from the outside. Only if the cognitive artifact is in control of executing its structures rather than needing to cope with the computational costs of processing the flood of perceptual stimuli can it, like natural cognitive systems, effectively control its input.

Therefore, it would be useless to let autonomous cognitive machines data-mine the infinitely rich structures of the world. What we consider a regularity, a law, a rule or simply a heuristics for decision-making is an arbitrary component of the data set that only makes sense in the light of our thinking or theorizing, and is therefore an anticipatory construction based on aspects of the cognitive system's past experiences: "Things are a construction of ours, the function of which is to emphasize the resemblance between aspects of our immediate experience and aspects of our past experience." [7]

## References

1. Angeline, P.J., Pollack, J.B.: Coevolving high-level representations. In: Langton, C., (ed.):Artificial life III. Addison-Wesley, Reading, pp. 55–71 ( 1994)
2. Arbib, M.: Schema theory. In: Shapiro, S. (ed.) Encyclopedia of artificial intelligence, 2nd edn. vol. 2, pp. 1427–1443. Wiley, New York (1992)
3. Arthur, W.B.: Inductive reasoning and bounded rationality. American Economic Review 84, 406–411 (1994)
4. Ashby, W.R.: Design for a brain. Chapman & Hall, London (1952)
5. Ashby, W.R.: An introduction to cybernetics, 2nd edn. Chapman & Hall, London (1956)
6. Bickhard, M.H.: Function, anticipation and representation. In: Dubois, D.M. (ed.): Computing anticipatory systems (CASYS 2000). American Institute of Physics, Melville, pp. 459–469 (2001)
7. Bridgman, P.W.: The nature of physical theory. John Wiley & Sons, New York (1936)
8. Brugger, P.: From haunted brain to haunted science: A cognitive neuroscience view of paranormal and pseudoscientific thought. In: Houran, J., Lange, R. (eds.): Hauntings and poltergeists: Multidisciplinary perspective. McFarland, Jefferson, pp. 195–213 ( 2001)
9. Cannon, W.B.: The wisdom of the body. Norton, New Yorks (1932)
10. Carnap, R.: Der logische Aufbau der Welt. Felix Meiner Verlag, Leipzig (1928). English translation: Carnap, R.: The logical structure of the world. Pseudoproblems in philosophy. University of California: Berkeley (1967)

11. Clancey, W.J.: Review of Rosenfield's The Invention of Memory. Artificial Intelligence 50, 241–284 (1991)
12. Clancey, W.J.: "Situated " means coordinating without deliberation. McDonnel Foundation Conference, Santa Fe (1992)
13. Conrad, K.: Die beginnende Schizophrenie. Versuch einer Gestaltanalyse des Wahns. Thieme, Stuttgart (1958)
14. Dennett, D.C.: Cognitive wheels: The frame problem of AI. In: Hookway, C. (ed.) Minds, machines, and evolution: Philosophical studies, pp. 129–151. Cambridge University Press, London (1984)
15. Dennett, D.C.: Consciousness explained. Little, Brown & Co, London (1991)
16. Diettrich, D.: A Physical Approach to the Construction of Cognition and to Cognitive Evolution. Foundations of Science 6, 273–341 (2001)
17. Drescher, G.L.: Made-up minds: A constructivist approach to artificial intelligence. MIT Press, Cambridge (1991)
18. Duhem, P.: The aim and structure of physical theory (French original published in 1906). Princeton University Press, Princeton (1954)
19. Duncker, K.: Zur Psychologie des produktiven Denkens. Springer, Berlin (1935). English translation: Duncker, K.: On Problem Solving. Psychological Monographs 58, 1–112 (1945)
20. Einstein, D., Menzies, R.: The presence of magical thinking in obsessive compulsive disorder. Behaviour Research and Therapy 42, 539–549 (2004)
21. von Foerster, H.: Molecular ethology. An immodest proposal for semantic clarification. In: Ungar, G. (ed.): Molecular mechanisms in memory and learning, pp. 213–248. Plenum Press, New York (1970) Reprinted in von Foerster, H.: Observing systems. Intersystems Publications, Seaside, pp. 149–188 (1982)
22. von Foerster, H.: Ethics and second-order cybernetics. Cybernetics & Human Knowing 1, 9–19 (1992)
23. Frost, R., Krause, M., McMahon, M., Peppe, J., Evans, M., McPhee, A., Holden, M.: Compulsivity and superstitiousness. Behaviour Research and Therapy 31, 423–426 (1993)
24. Gigerenzer, G.: Adaptive thinking. Oxford University Press, Oxford (2000)
25. von Glasersfeld, E.: Radical constructivism. Falmer Press, London (1995)
26. Gosselin, F., Schyns, P.G.: Superstitious perceptions reveal properties of internal representations. Psychological Science 14, 505–509 (2003)
27. Hesslow, G.: Conscious thought as simulation of behaviour and perception. Trends in Cognitive Sciences 6, 242–247 (2002)
28. Hong, F.T.: Deciphering the enigma of human creativity: Can a digital computer think? IPSP-2003 VIP Forum, October 4-11, 2003, Sveti Stefan, Montenegro (2003)
29. Hoyningen-Huene, P.: The nature of science. Nature & Resources 35, 4–8 (1999)
30. Huettel, S.A., Mack, P.B., McCarthy, G.: Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. Nature Neuroscience 5, 485–490 (2002)
31. Humphrys, M.: Action selection methods using reinforcement learning. PhD Thesis Trinity Hall, Cambridge (1997)
32. Kant, I.: Kritik der reinen Vernunft. Zweite Ausgabe. Reclam jun, Leipzig, English translation: Critique of Pure Reason, Second edition.(1781)
33. Keil, F.C.: Folkscience: Coarse interpretations of a complex reality. Trends in Cognitive Sciences 7, 368–373 (2003)
34. Kozhamthadam, J.: The discovery of Kepler's Laws. University of Notre Dame Press, Notre Dame (1994)

35. Lakatos, I.: Falsification and the methodology of scientific research programmes. In: Lakatos, I., Musgrave, A. (eds.) Criticism and the growth of knowledge, pp. 91–195. Cambridge University Press, London (1970)
36. Langer, E.J.: The illusion of control. Journal of Personality and Social Psychology 32, 311–328 (1975)
37. Lindeman, M., Aarnio, K.: Superstitious, magical, and paranormal beliefs: An integrative model. Journal of Research in Personality (in press)
38. Lorenz, K.Z., Tinbergen, K.: Taxis und Instinkthandlung in der Eirollbewegung der Graugans. Zeitschrift für Tierpsychologie 2, 1–29 (1939)
39. Lubow, R.E., Gewirtz, J.C.: Latent inhibition in humans: Data, theory, and implications for schizophrenia. Psychological Bulletin 117, 87–103 (1995)
40. Luchins, A.S.: Mechanization in Problem Solving. The Effect of Einstellung. Psychological Monographs 54/248 (1942)
41. Mach, E.: Knowledge and error. Sketches on the psychology of enquiry. (German original was published in 1905). Reidel, Dordrecht (1976).
42. Mach, E.: Popular scientific lectures. The Open Court, La Salle (Originally published in 1893) (1986)
43. Malinowski, B.: Magic, science and religion and other essays Free Press, Glencoe (Originally published in 1925) (1948)
44. Maturana, H.R.: Autopoiesis: reproduction, heredity and evolution. In: Zeleny, M. (ed.) Autopoiesis, dissipative structures and spontaneous social orders, pp. 48–80. Westview Press, Boulder (1980)
45. McAllister, J.W.: Phenomena and Patterns in Data Sets. Erkenntnis 47, 217–228 (1997)
46. McAllister, J.W.: The amorphousness of the world. In: Cachro, J., Kijania-Placek, K. (eds.): IUHPS 11th International Congress of Logic, Methodology and Philosophy of Science. Jagiellonian University: Cracow, 189 ( 1999)
47. McAllister, J.W.: Algorithmic randomness in empirical data. Studies in the History and Philosophy of Science 34, 633–646 (2003)
48. Neisser, U.: Cognitive psychology. Meredith, New York (1967)
49. Neisser, U.: Cognition and reality. W. H. Freeman, San Francisco (1976)
50. O'Regan, J.K.: Solving the " real " mysteries of visual perception: The world as an outside memory. Canadian Journal of Psychology 46, 461–488 (1992)
51. Oyama, S.: The ontogeny of information: Developmental systems and evolution (Republished in 2000). Cambridge University Press, Cambridge (1985)
52. Pörksen, B.: The certainty of uncertainty.(German original appeared in 2001). Imprint, Exeter (2004).
53. Porr, B., Wörgötter, F.: Inside embodiment. What means embodiment to radical constructivists? Kybernetes 34, 105–117 (2005)
54. Powers, W.T.: Behavior. The control of perception. Aldine de Gruyter, New York (1973)
55. Quine, W.V.: Word and object. MIT Press, Cambridge (1960)
56. Riegler, A.: When is a cognitive system embodied? Cognitive Systems Research 3, 339–348 (2002)
57. Rudski, J.M.: The illusion of control, superstitious belief, and optimism. Current Psychology 22, 306–315 (2004)
58. Sacks, O.: An anthropologist on Mars. Alfred A. Knopf, New York (1995)
59. Simon, H.A.: The architecture of complexity. In: Simon, H.A. (ed.) The sciences of the artificial, pp. 192–229. MIT Press, Cambridge (1969)
60. Simon, H.A.: Does scientific discovery have a logic? Philosophy of Science 40, 471–480 (1973)

61. Skinner, B.F.: ‘Superstition’ in the pigeon. Journal of Experimental Psychology 38, 168–172 (1948)
62. Spelke, E.S.: Core knowledge. American Psychologist 55, 1233–1243 (2000)
63. Tsang, E.W.K.: Superstition and decision-making. Contradiction or complement? Academy of Management Executive 18, 92–104 (2004)
64. Tyrrell, T.: Computational mechanisms for action selection. PhD thesis, University of Edinburgh, Centre for Cognitive Science (1993)
65. Vyse, S.A.: Believing in magic. In: The psychology of superstition, Oxford University Press, New York (1997)
66. Zebb, B.J., Moore, M.C.: Superstitiousness and perceived anxiety control as predictors of psychological distress. Anxiety Disorders 17, 115–130 (2003)