Strangeness

Sashank Varma, Dan L. Schwartz

Stanford Center For Innovations in Learning
450 Serra Mall, Building 160
Stanford, CA 94305-2055
sashank@stanford.edu

W. Lawrence Neeley Jr.

Center for Design Research, Mechanical Engineering Dept.
440 Escondido Mall, Building 530
Stanford, CA 94305-3030
wlneeley@stanford.edu

Abstract

Some systems are *strange*, near-bugs that turn out upon deeper analysis to be absolutely fundamental. This paper develops the notion of strangeness with respect to one class of AI system, *cognitive architectures*. It describes the detection of strange new architectures and their synthesis with conventional architectures. It also inventories strategies for the deliberate design of strange new architectures. Claims are justified by examples drawn from the history of (cognitive) science.

Introduction

Sometimes a system appears to have a bug, but upon further examination, proves to be working correctly – just not as expected. We call such systems (or such features of systems) *strange*.

This paper is about strangeness. The first half focuses on the process of detecting and acclimating to strangeness. For example, consider the discovery of irrational numbers by Hippasus. For this, he was drowned by his teacher Pythagoras, who could not "square" them with the other known numbers. Mathematicians did not acclimate to imaginary numbers until the eighteenth century, when de Moivre, Euler, and others established their mathematical utility. The detection of and acclimation to the strange is a general phenomenon that is not limited to technical disciplines. For example, Gertrude Stein remarked of modern art that it "looks strange and it looks strange and it looks very strange; and then suddenly it doesn't look strange at all and you can't understand what made it look strange in the first place."

The second half of the paper inventories a number of strategies for the deliberate design of strange systems. One strategy, which we will examine more closely below, is to emphasize one aspect of a conventional system to an incredible degree, to see how much it can carry. An example of this is APL, the first programming language designed for expressing mathematics. Dijkstra calls APL "a mistake, carried through to perfection." Once again, the deliberate design of the strange is not limited to technical disciplines. For example, van Gogh remarked of his painting process that "[i]nstead of trying to reproduce exactly what I see before my eyes, I use color more arbitrarily, in order to express myself forcibly."

To make the discussion tractable – our examples have already wandered over mathematics, computer science, and art – we focus on one class of system, *cognitive architectures*. These are AI formalisms for expressing computational models of cognition. Claims are justified primarily by anecdotes from the history of cognitive science (with Allen Newell playing a particularly prominent role), mathematics, and science. We conclude by introducing a metaphor for visualizing strangeness and calling for the formalization and expansion of this metaphor in future research.

Detecting and Acclimating to Strangeness

Consider the appearance of a strange new cognitive architecture containing a strange new computational mechanism, one that differs radically from the conventional mechanisms of conventional architectures. The strange mechanism will stand out against the field of normalcy (von Restorff 1933). Researchers will dishabituate to and take notice of it (Baillargeon, 1994). Its presence will unsettle them, causing internal conflict (Bruner & Postman 1949; Festinger, 1957; Petitto & Dunbar, 2004). The tension felt by individual researchers will grow to envelope the entire field (Kuhn 1977). Is the strange mechanism a bug? Or is it something more – the leading edge of a new paradigm?

There are two ways to relieve this tension. The logical way is to view the new mechanism as an error, a "contradiction" of a conventional mechanism, and to conduct research that selects between the two. The expectation, of course, is that the newcomer will be falsified and the conventional mechanism corroborated (Popper, 1963). The second, Hegelian response is to accept that both the conventional mechanism (the *thesis*) and the strange new mechanism (the *antithesis*) have value and to find a *synthesis* that unifies them while preserving each as a special case. It is in this way that strangeness spurs scientific progress (McAllister, 1996).

Dissonance

The first problem is deciding whether a new mechanism is an error, and should therefore be falsified, or whether it is strange, and should therefore be synthesized with a conventional mechanism. Because the former, logical response is usually appropriate, on those rare occasions when a strange new mechanism appears, problems will arise. Most researchers will dismiss it as "unscientific" and pay it no heed; others will attempt to falsify it, and when they fail, will brand it as "anomalous" (Barber, 1961; Kuhn 1996; Mahoney, 1977; Mitroff, 1974; Peters & Ceci, 1982). For example, when Stephen Grossberg's dissertation on artificial neural networks first appeared at Stanford in the mid 1960s, David Rumelhart remembers that:

> [W]e spent a great deal of time and effort trying to figure it out and failed completely, I should say, to understand what it was that Grossberg had actually done. […] There were those who thought that, well, 'There is something very deep here, and it's beyond us. We just can't figure it out.' And there were those who thought that this was just a story, and it meant nothing, and we shouldn't pay any attention to it. (quoted by Anderson & Rosenfeld, 1998, p. 271)

Ten years later, when Grossberg was an established scientist, Paul Werbos showed him the backpropagation algorithm, which he had just developed for his dissertation. Ironically, Grossberg's response was to dismiss it.

> The bottom line is, this stuff you've done, it's already been done before. Or else it's wrong. I'm not sure which of the two, but I know it's one of the two. (Werbos quoted by Anderson & Rosenfeld, 1998, p. 343)

These incidents illustrate the extent of the dissonance that strange new mechanisms cause.

However, some researchers will immediately recognize the value of a strange mechanism. For example, Newell describes his first exposure to Pandemonium in 1955 as a conversion experience:

> Oliver [Selfridge] had developed a mechanism that was so much richer than any other mechanism that I'd been exposed to that we'd entered another world as far as our ability to conceptualize. And that turned my life. I mean that was the point at which I started working on artificial intelligence. Very clear – it all happened one afternoon. (quoted by McCorduck, 1979, p. 134)

Synthesis

When a strange new mechanism is recognized and granted scientific standing, the work has only begun for now it must by synthesized with conventional mechanisms. This is a difficult and creative process. Researchers "must live in a world out of joint" (Kuhn 1996), attempting to reconcile seemingly contradictory mechanisms. Eventually, the surface differences will fade and a higher-order principle that subsumes them will emerge. Returning to the previous example, after learning about Pandemonium, Newell and Simon set about reconciling its parallel pattern-matching control structure with their own mechanisms for heuristic problem solving (Newell & Simon, 1972, p. 992; Simon, 1993, p. 255). The result was a series of strange new architectures culminating in the development of production system interpreters.

Another example comes from DARPA's funding of multiple speech understanding systems in the 1970s. Only two systems, Hearsay and Harpy, came close to meeting the program's ambitious goals (Cole 1980). Hearsay was an incremental advance on conventional production system architectures of the time. Harpy, by contrast, was a strange new architecture – what we might today call a probabilistic network. Newell was as an advisor on the Hearsay project. Although it would have been tempting to dismiss Harpy's success as anomalous, he did not. Instead, he attempted a Hegelian synthesis of the two systems. The result was the HPSA77 architecture, which matches Hearsay–like productions using a Harpy-like network (Newell 1980). This synthesis yielded new insights into cognition, such as the interpretation of working memory limitations as bottlenecks in the application of procedural knowledge.

Designing Strangeness

Strangeness is not just discovered; it can also be deliberately designed. In this sense, it differs from the serendipitous bugs that led Curie to radium and Fleming to penicillin. In this section, we describe a number of strategies for designing strange new cognitive architectures and computational mechanisms.

Empirical

Anomalous data are one source of strangeness. Although most experiments in cognitive science are Popperian exercises, a small number reveal aspects of cognition – unsuspected capacities and limitations – that cannot be explained by conventional architectures. When "[c]onfronted with an anomaly," the courageous researcher will be willing "to try anything," and will "transition from normal to extraordinary research" to do so (Kuhn 1996, p. 91).

One example of how anomalous data can spur the design of strange new architectures is the demonstrations of Bransford and his colleagues (e.g., Bransford & Johnson, 1973) of the *constructive* nature of language comprehension – that the meaning of an utterance is jointly constructed from the meanings of its constituent words and from the reader's schematic knowledge. These data were inconsistent with the then-dominant *interpretive* view, which focuses exclusively on word meanings. The Bransford data were initially dismissed, but eventually proved too compelling to ignore.

> What was interesting is the extent to which people picked up on it. They didn't believe it at first; then they were run as subjects in it, saw what happened, and said, My God, how could I ever have been so stupid as to not believe this? (Walter Weimer quoted by Baars, 1986, p. 306)

These anomalous data demanded the development of strange new architectures, such as the script systems of

Schank and Abelson (1977), that focus on the representation of schematic knowledge and its application during comprehension (and cognition more generally).

Another example where data spurred the design of strange new architectures dates back to the beginning of the cognitive revolution. Newell and colleagues developed the IPL family of programming languages to capture two fundamental characteristics of human cognition: the flexibility of problem solving and the associative nature of the memory (Newell, 1963, p. 87). These characteristics did not find natural expression in assembly language and Fortran, the programming languages of the time. To capture the first characteristic, they designed a new mechanism, dynamic memory allocation, that enables the construction of new data structures at runtime. This mechanism was necessary for implementing the LT, GPS, and NSS Chess programs, which progressively articulate problem spaces through search. To capture the second characteristic, they designed another mechanism: the linked list data structure and attendant functions for accessing it associatively, by content rather than by address. This mechanism was necessary for implementing the discrimination network at the heart of the EPAM model of verbal list learning (Simon, 1998).

Anomalous data have also spurred the design of strangeness in other fields. One example from mathematics is Weierstrass's discovery of an everywhere-continuous-but-nowhere-differentiable function: This pathological object prompted a reformulation of differentiation based on the strange new notion of "limit", which replaced older notions such as Newton's "fluxion" and Leibniz's "infinitesimal". Another example comes from chemistry, specifically "inversions" in the original periodic table (Kragh, 2000). Mendeleev's original periodic law was predicated on the *atomic weights* of elements. Although it worked well for most elements, there were a few anomalies, such as the fact that tellurium comes before iodine based on its chemical properties even though it has the higher atomic weight. This mystery was solved with Soddy's discovery in 1910 that the same element could exist in different isotopes and Moseley's demonstration in 1914 that the X-ray spectra of different isotopes of the same element possessed the same frequency, and that this frequency increased by a constant amount between consecutive elements in the periodic table. This physical property of an element, which came to be called its *atomic number*, was soon recognized as a better architectural principle for organizing the periodic table, and replaced atomic weight. (This change correctly orders tellurium and iodine.)

Methodological

A related source of strangeness is the development of new experimental methods and apparati. These enable the collection of new classes of data, data that usually escape conventional architectures and prompt the development of strange new architectures in which they are commensurate with existing data.

For example, consider the development of protocol analysis. In this experimental method, subjects "talk aloud" as they perform a complex cognitive task such as problem solving and discourse comprehension. As Newell remembers:

> As soon as we got the protocols [of thinking studies run at RAND in 1955-6] they were fabulously interesting. They caught and just laid out a whole bunch of processes that were going on. (Newell quoted by McCorduck, 1979, p. 212)

Protocol analysis differed from other methodologies of the time, such as the rapid presentation of stimuli using tachistoscopes, which were more appropriate for simple forms of cognition such as visual perception and memory retrieval. The resulting protocol data demanded new architectural accounts. In fact, they rather directly suggested what these accounts should be.

> My recollection is that I just sort of drew GPS right out of subject 4 on problem D1 – all the mechanisms that show up in the book, the means-end analysis, and so on. (p. 212)

The resulting computational models were far stranger than the statistical and mathematical models that dominated cognitive psychology at the time.

A more recent example is the development of functional magnetic resonance imaging, which has produced an explosion of data on the brain bases of cognition. These data demand new architectural accounts. Some are taking the conservative route of adapting existing architectures to this task. For example, Anderson, Bothell, Byrne, Douglass, Lebiere, and Qin (2004) have decomposed the ACT-R production system architecture into functional components (i.e., declarative memory, matcher, goal stack) and mapped each component to a different brain area (i.e., the hippocampus, striatum, and dorsolateral prefrontal cortex, respectively). Others are taking the more radical route of developing strange new architectures that place behavioral and neuroimaging data on equal footing. For example, the 4CAPS architecture (Just & Varma, 2005) casts cognition as the emergent product of collaborative processing among a confederacy of brain areas, each modeled as an encapsulated production system with its own knowledge sources and resources for fueling computation.

Theoretical

Strangeness can also be generated from within, without pressure from external sources such as anomalous data and new experimental methods. Doing so requires researchers to adopt the persona of designers and to employ methods that encourage exploration and promote divergent thinking. One such method, *synectics*, was invented by William Gordon (1961). One move of synectics is to make the familiar strange; another is to make the strange familiar. We consider three examples of the former and one of the latter.

Emphasis. One strategy for making the familiar strange is to emphasize one aspect of cognition to an incredible degree and to see how much of cognition it can carry. Sometimes, the answer is more than one would otherwise suspect, the result strange accounts of seemingly very different aspects of cognition. One example is Newell's ongoing attempt, beginning with GPS and continuing through Soar, to construe all of cognition as problem solving. The result has been a number of unexpected successes, such as Miller and Laird's (1996) model of categorization, which operationalizes concepts as collections of exemplars, each a chunked production acquired during a previous categorization episode. That exemplar-based categorization can be treated as an instance of problem solving is strange.

A variant of emphasis is used when an architecture proves problematic in some way. The natural response is to add a new mechanism that directly solves the problem. Another approach is to use emphasis: to look *hard* at the current repertoire of mechanisms and to generalize and compose them to solve the problem in an unexpected way. The result can transform a conventional architecture into a strange one. For example, the development of large production system models in the 1970s raised the problem of conflict resolution, or how to select which of the multiple matching productions to fire next. A number of new mechanisms were proposed to solve the problem, each implementing a different conflict resolution scheme (i.e., principled policy for making selections). For example, the MEA scheme of the OPS5 architecture selects in a way that produces means-ends problem solving. However, none of the proposed schemes proved to be sufficiently general. Newell was of the opinion that when a problem arises, "if you look at the architecture hard enough it will tell you how to solve that problem" (quoted by Agre, 1993, p. 447). Soar therefore made the strange decision to *not* include a conflict resolution scheme (Newell, 1990). Instead, it emphasizes conventional mechanisms, combining productions and declarative elements in a recursive manner to select which of the matching productions to fire next.

Researchers in other disciplines have also used emphasis to generate strange solutions to problems. For example, Jack Cowan, a pioneer in artificial neural networks, describes the following exchange between the physicist Paul Dirac and a graduate student:

> Then another student asked him, 'How did you discover antimatter?' Dirac said, 'Oh, that was easy. In relativity, energy's the square of a quantity, so I just took the square root.' (p. 124 of Anderson & Rosenfeld, 1998)

What could be stranger than antimatter? As Steven Weinberg has observed, "This is often the way it is in physics – our mistake is not that that we take our theories too seriously, but that we do not take them seriously enough."

Negation. Another strategy for making the familiar strange is to take a conventional architecture and to negate one of its assumptions. The result is typically internally inconsistent, but sometimes it is a strange new architecture. This is an example of Feyerabend's (1988, p. 14) dictum that "given any rule, however "fundamental' or 'rational', there are always circumstances when it is advisable not only to ignore the rule, but to adopt its opposite." For example, consider again the problem of conflict resolution in production system architectures. As we just saw, most architectures solve this problem by including a mechanism – a conflict resolution scheme – for selecting which matching production to fire next, and Soar uses the emphasis strategy to solve this problem in a strange way. By contrast, 4CAPS (Just & Varma, 2005) and EPIC (Meyer & Kieras, 1997) have adopted the negation strategy, denying that a conflict resolution scheme is necessary at all. Instead, they simply fire all matching productions in parallel. (To prevent combinatorial explosion, 4CAPS constrains the supply of computational resources whereas EPIC relies on bottlenecks on perceptual inputs into and motor outputs out of the system.) In this regard, these architectures are strange.

The canonical use of the negation strategy to generate strangeness is the formulation of non-Euclidean geometry: For centuries, mathematicians attempted to deduce the fifth of Euclid's postulates – for a given line and point not on that line, there is exactly one line parallel to the given line that passes through the given point – from the other four postulates, which seem simpler by comparison. Around 1800, mathematicians began asking a different question: If the fifth postulate is negated, is the result still a valid geometry? The answer, surprisingly, is "yes". The result: a number of non-Euclidean geometries, including the Riemannian geometry that undergirds Einstein's general theory of relativity.

Cross-Pollination. A third strategy for making the familiar strange is cross-pollination: When the mechanisms of seemingly inconsistent architectures are juxtaposed, the result is sometimes a strange new architecture. For example, there were a number of efforts in the 1980s to combine the mechanisms of symbolic and connectionist architectures, efforts undertaken by researchers well-versed in the both paradigms. For example, Rumelhart says of the development of localist connectionist architectures:

> I was also inspired by the work of Ross Quillian, who in those days was doing computer models of so-called semantic networks. (quoted on p. 272 of Anderson & Rosenfeld, 1998)

He names another symbolic influence on this work:

> I had a student named Jim Levin. He got interested in a system that he called Proteus. Proteus was inspired by Carl Hewitt's Actor system, but it turned out to be as close as anything to neural networks. (p. 273)

The result of these cross-pollinations was a number of strange new architectures.

This brings up an interesting question: Going forward, which cross-pollinations are most likely to yield strange new architectures? In our opinion, one promising cross-pollination is combining the mechanisms of symbolic architectures and exemplar memories. By symbolic architectures, we mean production system architectures, whose mechanisms include productions and declarative memories. By exemplar memories, we mean systems that contain large numbers of traces, each colored by the context in which it was encoded, that operate collectively; an example is Minerva-II (Hintzman, 1986). There have been a handful of attempts to cross-pollinate symbolic and exemplar architectures over the past two decades. For example, early versions of ACT-R (Anderson, 1993) and the Construction-Integration model (Kintsch, 1988) employ symbolic mechanisms for immediate processing and exemplar mechanisms for prior knowledge. Another approach is exemplified by Miller and Laird's (1996) Soar model of categorization and Stanfill and Waltz's (1986) massively parallel memory: these employ only symbolic mechanisms, but of a less abstract and more exemplar nature than is typical, and that operate by mass action. We are optimistic that a new round of cross-pollination between the symbolic and exemplar paradigms will yield strange new architectures.

Importation. In addition to making the familiar strange, we can also make the strange familiar. There exist novel computational formalisms in other technical disciplines. Although most have nothing to offer AI, there are exceptions. These formalisms must be identified and imported into AI, where they can serve as strange architectures.

In fact, there is a long history of the formalisms of other disciplines finding their way into AI. One early example is Norbert Wiener's importation of information theory and control theory from engineering into AI; these formalisms provided strange new accounts of language processing, decision making, and other aspects of cognition. Another example is the production system formalism, which was originally proposed by the logician Emil Post as a mathematical theory of computation on par with Turing machines.

> The production system was one of those happy events, though in a minor key, that historians of science often talk about: a rather well-prepared formalism, sitting in wait for a scientific mission. (Newell & Simon, 1972, p. 889)

Newell and Simon imported this formalism into AI, where it provided a strange new account of problem solving. At about the same time, Chomsky imported it into linguistics, producing a new architecture, generative grammar, for language. A final example is the importation of statistical mechanics and the theory of spin glasses into AI during the 1980s. These physical formalisms cross-pollinated with conventional connectionist mechanisms. The result was strange new architectures such as Hopfield (1982) networks and Boltzmann machines (Ackley, Hinton, & Sejnowski, 1985).

Institutional

The fourth strategy for developing strange new architectures is through institutional planning. This is not as surprising as it might seem at first glance. Consider that brainstorming was invented by Alex Osborn as a solution to the lack of creativity he consistently encountered in meetings. Brainstorming establishes a safe environment in which participants with different talents feel free to offer strange new ideas. The planning of AI institutions where new data, new methods, and new computational formalisms mix freely is just brainstorming writ large. The expected outcome is stranger architectures that are produced by comparatively homogeneous institutions. For example, during the early 1980s, CMU's School of Computer Science was still a hotbed of symbolic AI. Geoffrey Hinton was hired away from UCSD because the faculty were interested in "getting a neural network presence" (Hinton quoted by Anderson & Rosenfeld 1998, p. 375). This plan succeeded: Hinton and his symbolic colleagues collaborated on a number of strange new architectures that combine connectionist and symbolic mechanisms (e.g. Touretzky & Hinton, 1988).

Cross-pollination and institutional planning are in some sense the same strategy applied at different levels. Cross-pollination operates at the individual level, requiring researchers who have been trained in multiple architectural paradigms. Such researchers are few and far between. Institutional planning acknowledges this state of affairs, applying cross-pollination at a broader social level: although each researcher might know only a single architecture, by working together in the same organization, they can cross-pollinate familiar mechanisms to produce strange new architectures.

Conclusion

The fundamental claim of this symposium – that some bugs are informative – coheres well with existing research programs in AI, such as automated scientific discovery (e.g., Shrager & Langley, 1990) and model-based diagnosis (e.g., Ginsberg, 1987). We make a different, though related claim: some systems are strange, near-bugs that turn out upon deeper analysis to be absolutely fundamental. This paper has developed the notion of strangeness with respect to one class of system in AI, cognitive architectures (and the computational mechanisms that compose them). It has described the detection of strange new architectures and their synthesis with conventional architectures. Finally, it has inventoried strategies for the deliberate design of strange new architectures.

An important question is whether AI formalisms can be developed for representing, reasoning over, and reconciling seemingly contradictory architectures and mechanisms? Recall that the normal response to contradiction is logical: to retract one of the conflicting alternatives. This is the kind of move for which truth maintenance systems (e.g., Ginsberg, 1987) are well-suited. However, strangeness demands a different move: the conflicting alternatives must be reconciled in a new synthesis. This requires AI formalisms that tolerate contradictions without falling into tautology; that afford their synthetic resolution; and that support the design strategies described above.

Interestingly, Wittgenstein (1930) anticipated the need for such formalisms.

> I predict a time when there will be mathematical investigations of calculi containing contradictions, and people will actually be proud of having emancipated themselves from consistence.

An important goal for future research, then, is to develop such formalisms.

We believe that formalisms capable of handling the strange will be statistical and inductive, not logical and deductive, in nature. For example, consider the following metaphor for visualizing the synthesis of the conventional and the strange: In simple linear regression, the goal is to find a line that best accounts for (i.e., minimizes the variance in) a set of points. An example is shown in Figure 1a. We can interpret this diagram at different levels of abstraction. For example, the points can be data on human cognition, the line a mechanism that predicts them, and the axes different architectures. Another possibility is that the points are mechanisms, the line an architecture, and the axes different architectural paradigms. Regardless of the interpretation chosen, interpret Figure 1a as the current understanding of the field. Next, consider the appearance of a new point, as shown in Figure 1b. The point is an outlier. A well-known problem in regression is how to handle outliers. The outlier – a strange new datum or mechanism – introduces tension into the field. One way to alleviate this tension is to classify it as an error or a member of a different distribution, and to ignore it. This is the logical response. It contrasts with the Hegelian response, which is to reconcile it with the other points. This maps in the regression metaphor to assuming that, appearances to the contrary, the outlier belongs to the same distribution as the other points. The next step is to sample additional points in the empty region and to estimate a new line that accounts for them all. The result, shown in Figure 1c, can be a strange account, one orthogonal to current understanding. This represents a Hegelian synthesis of the conventional and the strange.

This metaphor nicely captures the synthesis of the strange and the conventional. Can it be formalized? Can it be stretched to accommodate the strategies for designing strange new architectures inventoried above? We believe these questions are worth pursuing. Critically, the design of new architectures does not affect the contents of the space – the points and lines – but rather the space itself, specifically the identity of its axes. For example, the empirical and methodological strategies can be viewed as rotating the space so that an outlier (i.e., anomalous datum) or set of outliers (i.e., class of anomalous data collected from a new method) relate coherently with the existing points. The theoretical and institutional strategies can be viewed as requiring the extraction of additional axes. For example, Principle Components Analysis is a multivariate technique for extracting a sufficient number of orthogonal axes for coherently plotting a set of points. It might be possible to use this technique and related statistical and machine learning methods to induce strange new architectures (i.e., axes) that better organize the data on cognition.

References

Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. 1985. A learning algorithm for Boltzmann machines. *Cognitive Science* 9: 147-169.

Agre, P. E. 1993. Interview with Allen Newell. *Artificial Intelligence* 59: 415-449.

Anderson, J. 1993. *Rules of the Mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. A., and Rosenfeld, E. eds. 1998. *Talking nets: An oral history of neural networks*. Cambridge, MA: MIT Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. 2004. An integrated theory of the mind. *Psychological Review* 111: 1036-1060.

Baars, B. J. 1986. *The cognitive revolution in psychology*. New York: The Guilford Press.

Baillargeon, R. 1994. How do infants learn about the physical world? *Current Directions in Psychological Science* 3: 133–140.

Barber, B. 1961. Resistance by scientists to scientific discovery. *Science* 134: 596-602.

Bransford, J.D., and Johnson, M.K. 1973. Considerations of some problems of comprehension. In W. Chase (ed.),

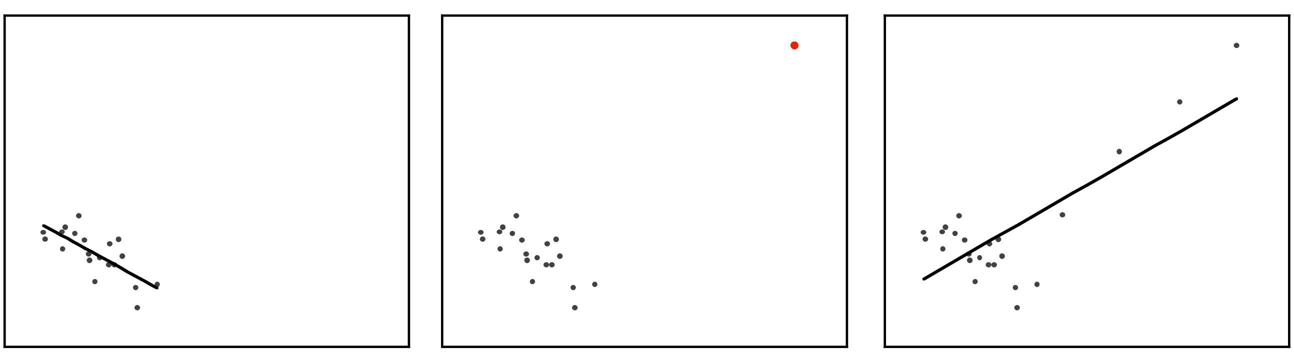

Figure 1. (a) Conventional understanding. (b) Outlier. (c) New synthetic explanation.

Visual information processing (pp. 383-438). New York: Academic Press.

Bruner, J. S. and Postman, L. 1949. On the perception of incongruity: A paradigm. *Journal of Personality* 18: 206-223.

Cole, R. ed. 1980. *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Festinger, L. 1957. *A theory of cognitive dissonance.* Evanston, IL: Row Peterson.

Feyerabend, P. 1988. *Against method revised edition*. London: Verso.

Ginsberg, M. L. ed. 1987. *Readings in nonmonotonic reasoning*. Morgan Kaufmann: San Mateo, CA.

Gordon, W. J. 1961. *Synectics*. New York: Harper and Row.

Hintzman, D. L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93: 411-428.

Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79: 2554-2558.

Just, M. A., and Varma, S. 2005. *The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition*. Manuscript submitted for publication.

Kintsch, W. 1988. The role of knowledge in discourse comprehension construction-integration model. *Psychological Review* 95: 163-182.

Kragh, H. 2000. Conceptual changes in chemistry: The notion of a chemical element, ca. 1900-1925. *Studies in History and Philosophy of Modern Physics* 31: 435-450.

Kuhn, T. S. 1977. *The essential tension. Selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Kuhn, T. S. 1996. *The structure of scientific revolutions*. 3rd ed. Chicago: The University of Chicago Press.

Mahoney, M. J. 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1: 161-175.

McAllister, J. W. 1996. *Beauty and revolution in science*. Ithaca, NY: Cornell University Press.

McCorduck, P. A. 1979. *Machines who think*. San Francisco: W. H. Freeman.

Meyer, D. E., and Kieras, D. E. 1997. A computational theory of executive control processes and human multiple-task performance: Part 1. Basic Mechanisms. *Psychological Review* 104: 3-65.

Miller, C. S., and Laird, J. E. 1996. Accounting for graded performance within a discrete search framework. *Cognitive Science* 20: 499-537.

Mitroff, I. I. 1974. Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. *American Sociological Review* 38: 579-595.

Newell, A. 1963. Documentation of IPL-V. *Communications of the Association of Computing Machinery* 6: 86-89.

Newell, A. 1980. Harpy, production systems, and human cognition. In R. Cole (ed.), *Perception and production of fluent speech* (pp. 289-395). Hillsdale, NJ: Lawrence Erlbaum.

Newell, A. 1990. *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., and Simon, H. A. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Peters, D. P., and Ceci, S. J. 1982. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5: 187-255.

Popper, K. R. 1963. *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper and Row.

Schank, Roger C., and Abelson, Robert P. 1977. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Shrager, J. and Langley, P. Eds. 1990. *Computational Models of Scientific Discovery and Theory Formation.* San Mateo, California: Morgan Kaufmann.

Simon, H. A. 1993. Allen Newell: The entry into complex information processing. *Artificial Intelligence* 59: 251-259.

Simon, H. A. 1998. Allen Newell: 1927-1992. *IEEE Annals of the History of Computing* 20: 63-76.

Stanfill, C. and Waltz, D. 1986. Toward memory-based reasoning. *Communications of the ACM* 29: 1213-1228.

Touretzky, D. S., and Hinton, G. E. 1988. A distributed connectionist production system. *Cognitive Science* 12: 423-466.

Von Restorff, H. 1933. Uber die Wirkung von Bereichsbildungen im Spurenfeld The effects of field formation in the trace field. *Psychologie Forschung* 18: 299-34.