Diagrams as Scaffolds for Abductive Insights

Michael H.G. Hoffmann

Georgia Institute of Technology
Philosophy Program, School of Public Policy
685 Cherry Street, N.W.
Atlanta, GA 30332-0345, USA

Abstract

Based on a typology of five basic forms of abduction, I propose a new definition of abductive insight that emphasizes in particular the inferential structure of a belief system that is able to explain a phenomenon after a new, abductively created component has been added to this system or the entire system has been abductively restructured. My thesis is, first, that the argumentative structure of the pursued problem solution guides abductive creativity and, second, that diagrammatic reasoning—if conceptualized according to the requirements defined by Charles Peirce—can support this guidance. This support is mainly possible based on the normative power of the system of representation that has to be used to construct diagrams and to perform experiments with them.

Introduction

The goal of abductive reasoning is to generate a hypothesis that can explain a surprising or unexplained fact (Peirce CP 5.171). If this hypothesis exists already in our mind or in a database, we can call it "selective abduction," and if we have to create a new hypothesis—either a historically new one or one that is new for a student who learns something by discovery—we can talk about "creative abduction" (Magnani 2009).

I want to focus here on two problems of creative abduction. The first one is the question of how exactly do we know that a newly created hypothesis "explains" something. For computer programs this question points to the need of a stopping rule; a rule that "tells" the program when exactly the goal of discovering something new has been achieved. In cognitive science this point of discovery is described as an "insight" or an "Aha! event." But what exactly is an "insight"? When can the search for a solution stop?

The second problem of abductive creativity I want to discuss is the question of how the process of discovering an explanatory hypothesis can be guided by diagrammatic reasoning. Research on insights shows that, on a cognitive level, the genesis of an insight is far less discontinuous than it is usually perceived. While we experience insights as coming out of the blue and genuinely surprising, Kenneth Bowers and his colleagues showed in a series of experiments that the conscious recognition of a problem solution is preceded by an unconscious "intuitive phase" in which subjects come gradually closer to this solution. Even in situations where the subjects could not identify a certain gestalt or concept, they were able to distinguish, at an above-chance level, between coherent and non-coherent stimuli. This indicates that even when there is no final Aha! event, important things are happening unconsciously (Bowers, Farvolden and Mermigis 1995). These results support the thesis that creativity in general should be understood as an incremental process rather than an event. This means that the question of how this process can be supported and facilitated becomes even more important. My thesis is that diagrammatic reasoning can play a crucial role as a "scaffold" for abductive creativity.

This thesis is hardly new when formulated in these general terms (Larkin and Simon 1987; Barwise and Etchemendy 1994; Cheng and Simon 1995; Hegarty 2004; Nersessian 2008). What is new, however, is the understanding of "diagrammatic reasoning" that I will propose. It is not based on the usual contrast of "diagrammatic" versus "sentential" representations, but on a definition of "diagram" that Charles Peirce developed about a hundred years ago. According to his approach, a diagram is, first of all, a representation of relations that is constructed by means of a "system of representation." Such a system is defined by a set of rules, conventions, and a certain ontology. Based on this definition, it is crucial for diagrammatic reasoning—defined as a process of constructing diagrams, experimenting with them, and observing the results of this experimentation—that the outcome of experiments with diagrams is determined by the rules of the chosen system of representation. Thus, diagrammatization systems exert a normative power which, I will argue, is essential for its function as a "scaffold" for abductive creativity.

I will start, however, with a typology of different forms of abductive creativity and three examples that provide the basis for my conceptualization of what an "insight" is.

Forms of Abduction

After Peirce introduced "abduction" as a third form of reasoning besides deduction and induction, many attempts have been undertaken to distinguish different types "of forming an explanatory hypothesis" (Peirce CP 5.171). Based on G. Schurz's (2008) suggestion to differentiate "patterns of abduction" according to the types of things that are abductively inferred—that is, the things that are essential for providing the explanation in an explanatory hypothesis—I would propose the following typology of abductive inferences:

1. Fact abduction: When a single fact provides an explanation. Schurz distinguishes here further between "observable" and "unobservable fact abduction." An example of the first type would be if the fingerprints on this glass can be explained by the fact that my friend Peter used it. Unobservable fact abduction is performed when we explain these huge footprints in the rock by the fact that a certain dinosaur went along what was originally a beach.
2. Type abduction: This footprint on the beach can be explained by the fact that a human being, not an animal, went here. Or: the concept of "inertia" can explain that the planets circle in predictable paths around the sun.
3. Law abduction: When a general observation like "all pineapples taste sweet" gets explained—based on the background knowledge "whatever contains sugar tastes sweet"—by the hypothesis "all pineapples contain sugar." Or when Boyle explains the behavior of gases in his experiments by the law that determines, in a closed system and when the temperature is kept constant, the relationship between the absolute pressure and volume of a gas as being proportional ($pV=k$).
4. Theoretic model abduction: When an explanation is only possible by creating a model that relates facts or types on the one hand and laws on the other. An example is Archimedes's explanation of the phenomenon that some objects are swimming on water while others are sinking by means of the buoyancy model: if the force exerted by a certain amount of water is greater than the force exerted by an object, then the body will swim; if smaller, it will sink. — There are two preconditions for theoretic model abduction. The first one is that we need a system of representation that provides the means necessary to represent a model, that is a language that provides a certain ontology, semantics, and syntax. The second condition is what Peirce called a "theoric step," or a "theoric transformation," that is the process of finding an adequate perspective on a problem so that we can decide which representational system is adequate to model the problem. (For Peirce, "theoric" refers to the Greek "theoria," whose original meaning is "vision"; see Hoffmann 2005; Kaplan and Simon 1990 talked about a "change in representation" that is necessary for the solution of certain problems, and Dominowski 1995 and Smith 1995 are using the term "restructuring" of a problem space). The theoric step in theoretic model abduction determines how we "frame" a problem, as we could say. Sometimes, there will be no need for any new fact, type, or law, but only a shift of perspective, a reframing or restructuring. And sometimes there might be a variety of different models that can explain a phenomenon equally well. The significance of theoric steps will become clear with the examples that I describe in the next section.
5. Meta-diagrammatic abduction: According to the understanding of diagrammatic reasoning that I will develop below, any construction of a theoretical model is a result of diagrammatic reasoning. Since diagrammatic reasoning—as already mentioned—presupposes a certain system of representation, it is clear that completely new theoretic models are possible when we change or develop those representation systems themselves. Exactly this is the function of "meta-diagrammatic abduction." An example would be the development of non-Euclidean geometries out of Euclid's geometry which provided the means for a whole new set of theorems and proofs.

There might be more types of things that can be abductively inferred, but these most general types are indispensable for a more precise understanding of what it means to create an "explanatory hypothesis." My argument for this minimal list can only be formulated after we see more clearly how an explanatory insight can be described. But the basic idea of this typology can already be determined. Explanations can be provided by particular facts (1), general types of entities or concepts (2), general laws (3), or by general theoretic models that combine these in different or new ways (4). Since model construction itself depends on the representational means available, there must be something like meta-general abduction that refers to the possibility of creating new representational systems (5).

After this distinction of different forms of creative abduction, let me turn now to the first problem mentioned above: The question of how we can know when an abduction is completed, a phenomenon "explained," or an "insight" achieved. For this, I will analyze in the following section three examples: the first one is an example for theoretic model abduction, the second one for model and type abduction, and the third one for meta-diagrammatic abduction.

Examples of Abductive Insights

Imagine you are an artist who wants to cut a large marble block into 27 equal cubes. Of course, you could do it by the six cuts numbered in Figure 1. However, it might also be possible to do the job with less than six cuts if the pieces are rearranged between each cut as indicated in Figure 2. So, the question is: Is it possible to do it with less than six cuts, and how could it be done?

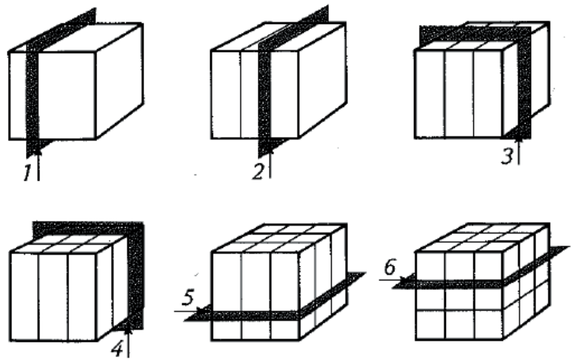

Figure 1: Cutting a marble block by six cuts (from: Jacobs 1970)

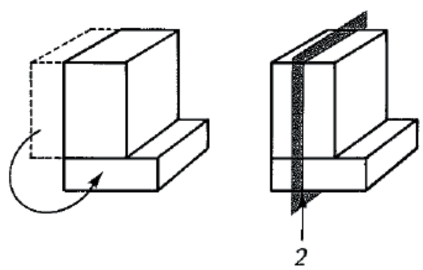

Figure 2: Rearranging the pieces to do it with less than six cuts

My second example concerns the problem of how to prove that the sum of a triangle's inner angles add up to 180°. As with my first example, it helps to visualize the problem by means of a diagram as the one in Figure 3.

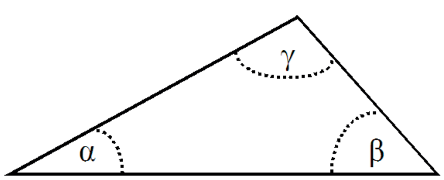

Figure 3: How to prove that the sum of the three angles is 180°?

The third example is taken from the history of environmental philosophy and describes Aldo Leopold's development of the notion "thinking like a mountain" as a theoretic model abduction that is based on a new representation system and, thus, an example for meta-diagrammatic abduction. Bryan Norton describes in his book *Sustainability. A Philosophy of Adaptive Ecosystem Management* how Leopold reframed over time his conceptualization of the role of human beings in nature. In his famous piece "Thinking Like a Mountain," Leopold describes how in his early days "we had never heard of passing up a chance to kill a wolf." The argument was simple: "I thought that because fewer wolves meant more deer, that no wolves would mean hunters' paradise" (Leopold 1949).

But seeing state after state extirpating its wolves, he realized that the following overpopulation of deer did not only change the balance of the ecosystem on the mountains, but the geology of the mountains themselves:

> Such a mountain looks as if someone had given God a new pruning shears, and forbidden Him all other exercise. … I now suspect that just as a deer herd lives in mortal fear of its wolves, so does a mountain live in mortal fear of its deer.

For Norton, the most central insight of Leopold's "is an insight about the importance of scale, both temporal and spatial, in our thinking about environmental management." Based on a distinction of "human, experiential time," "ecological time," and "geological time," he interprets Leopold's title as follows:

> Learning to think like a mountain is to expand one's temporal consciousness to see humans as actors not just on a short-term economic stage, but also as increasingly dominant actors on the ecological scale, capable of changing not just the actors on the stage but also the very stage itself. (Norton 2005, 214)

Such a shift of perspective—a "theoric step" in Peirce's words—is also the necessary precondition to solve the problem of the marble block. There is only one answer to the question whether it is possible to cut the cube with less than six cuts: No. The necessity of this answer becomes immediately evident if we shift our attention to the fact that, whatever we do regarding the rearrangement of the pieces, there is always one cube in the middle of the original block. Because this cube has six sides that need to be cut, there is no way to do the job with less than six cuts.

The solution of this problem is possible by what I introduced above as a "theoric step": we need a shift of perspective, that is we need to focus on the innermost cube in order to create an explanation of our answer. This new perspective does neither require the creation of a new fact, nor a new type, law, model, or representation system. We don't create any new entity; the only thing we create is a new perspective on the problem.

While the problem of the cube could also be resolved by simply trying all possible arrangements for cutting the pieces, there are others that are impossible to solve without a theoric transformation. Peirce hints at the proof of Desargues's theorem where it is necessary to perceive a two-dimensional constellation of triangles and lines as a three-dimensional pyramid that is cut by two planes (Hoffmann 2005).

A possible way to solve the triangle problem is to use an auxiliary line. As showed in Figure 4, we can introduce a parallel to the triangle's base through its apex. Based on the Euclidian axioms we know that $\alpha = \alpha'$ and $\beta = \beta'$, so that $\alpha+\beta+\gamma = \alpha'+\beta'+\gamma = 180°$.

This proof is what "explains" that the sum of the inner angles in a triangle equals 180°. Part of this explanation, however, is the auxiliary line which was nowhere mentioned in the description of the problem. To create the "new idea" of this line we need to perform what I called "type abduction." (It is not "observable fact abduction," because even though the auxiliary line is clearly visible in the diagram, Euclid's axiom do not apply to the tokens by which we visualize theorems, but only to the general types that are represented by the tokens—it is not the visible representation that proves anything, but only the deductive relation between general propositions.)

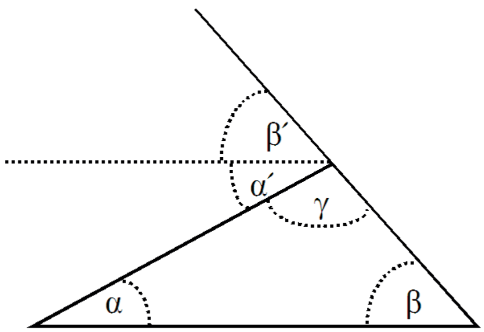

Figure 4: Proving the theorem about the inner angles of a triangle by means of an auxiliary line

Defining "Insight"

Although the solution of both these problems is immediately evident to us—if we did not know the answers, it is indeed an Aha! event that we experience—we should analyze this experience in some more detail to get a better understanding of what exactly an "insight" is. With regard to this question, John Clement suggested a distinction between "three categories of insight behavior" that he dubbed "breakthrough," "scientific insight," and "pure Eureka event":

> A breakthrough is a process that produces a key idea – an important component of a solution – and that overcomes a barrier that can block progress toward a solution.
>
> A scientific insight is a breakthrough occurring over a reasonably short period of time leading to a significant structural improvement in one's model of a phenomenon. That is, it constitutes a shift from the subject's previous way of representing the phenomenon and leads to an increase in understanding of the phenomenon …
>
> A pure Eureka event is a scientific insight where: (1) there is an extremely fast emergence of a new idea with little evidence of preparation; (2) the new idea is a whole structure replacing the subject's previous model or understanding of a situation; (3) the process is not explainable via normal reasoning processes; extraordinary thought processes that are unconscious or different from normal thought processes are involved. (Clement 2008, 103-104)

The concepts used in these definitions, however, are too vague to be operational. What exactly is "an important component of a solution"? What "a significant structural improvement in one's model"? An "increase in understanding"? What is "a reasonably short period of time" in contrast to "an extremely fast emergence of a new idea"? What is "normal reasoning"? The main problem in Clement's definitions, however, is the fact that it remains unclear how exactly a "solution" of a problem can be characterized. When can we stop looking for a solution? With regard to this question, Steven Smith's suggestion is helpful when he defines:

> Insight experiences occur when restructuring yields a knowledge state in which many, or perhaps all, of the important constraints or needs of a problem are suddenly satisfied. (Smith 1995, 143)

An even shorter definition has been proposed by John and Susan Josephson in their book *Abductive Inference*: "To understand something is to grasp an explanation of it." This definition, of course, presupposes an understanding of "explanation":

> An explanation is an assignment of causal responsibility; it tells a causal story (at least this is the sense of "explanation" relevant to abduction). Thus, finding possible explanations is finding possible causes of the thing to be explained, and so abduction is essentially a process of reasoning from effect to cause. (Josephson and Josephson 1994, 28-29)

Causality alone, however, seems too narrow for a comprehensive understanding of "explanation." Although both my marble block and the triangle example provide an explanation for a phenomenon, it would be a stretch to talk about a "causal story" in these cases. Nevertheless, leaving this point on explanation aside, both these approaches can be used as a starting point to formulate the following definition of "insight": An insight resulting from the creation of an explanatory hypothesis is the experience that what someone created in fact, type, law, model, or meta-diagrammatic abduction fits into a system of beliefs—or provides such a system—that fulfills three conditions: (1) each one of the beliefs in the system is acceptable to the person experiencing the insight; (2) the system as a whole satisfies for this person the constraints or needs given by the problem at hand; and (3) this system is explanatory in the sense that it could be represented in the form of an acceptable argument whose conclusion is a proposition that describes the phenomenon that needs to be explained.

Before I discuss some questionable points of this definition, let me illustrate its meaning by specifying for each of my examples the argument that would explain the phenomenon in question. To explain the fact that at least six cuts are necessary for the marble block, model abduction would create an explanatory system of beliefs that can be represented as follows:

- If the inner cube has six sides, then at least six cuts are necessary
- The inner cube has six sides
- Therefore, at least six cuts are necessary

Since each of the premises is acceptable, the problem solved, and the argument acceptable since it is deductively valid and explanatory, all three conditions for an insight are fulfilled. If one of the premises would not have been so easily acceptable, it would be possible to add further arguments for them to create eventually a more complex argumentation. In such a case, different forms of abduction would be necessary to create an insight.

Such a combination of several forms of abduction is necessary to solve the triangle task. What we need is, on the one hand, a model of the final proof and, on the other, the auxiliary line that can be created by means of type abduction. The coherent system of beliefs could then be represented as follows:

- If the dotted line in Figure 4 is parallel to the triangle's base, then $\alpha = \alpha'$ and $\beta = \beta'$
- If $\alpha = \alpha'$ and $\beta = \beta'$, then $\alpha + \beta + \gamma = \alpha' + \beta' + \gamma$
- $\alpha' + \beta' + \gamma = 180°$
- The dotted line is parallel to the triangle's base
- Therefore, $\alpha + \beta + \gamma = 180°$

Again, if someone has problems to accept one of the premises, further arguments would be possible to defend them.

The example of Leopold's change of mind is interesting because it illustrates the possibility that the same problem situation can be framed by competing systems of beliefs that lead to conflicting conclusions. Leopold's starting point can be reconstructed as follows:

- If you want "hunters' paradise," enlarge the deer population as far as possible
- If you want to enlarge the deer population as far as possible, kill as many wolves as possible
- You want "hunters' paradise"
- Therefore, kill as many wolves as possible.

An essential step of Leopold's shift of perspective can be reconstructed as the realization of the following argument:

- The wolf is the only predator of the deer
- We are killing as many wolves as possible
- If the wolf is the only predator of the deer, and if we are killing as many wolves as possible, then the deer population grows
- If the deer population grows, all the edible bushes, trees, and seedlings on the mountain will be destroyed
- If all the edible bushes, trees, and seedlings on the mountain will be destroyed, erosion sets in
- If erosion sets in, the mountain will be destroyed
- Therefore, the mountain will be destroyed

This argument can motivate the following argument whose conclusion contradicts that of the first one:

- You do not want to destroy the mountain (at least if you "think like a mountain")
- If you do not want to destroy the mountain, you should protect the wolves (supported by the previous argument)
- Therefore, you should protect the wolves

Based on the logical validity of these arguments, the initial and the final system of beliefs are obviously each in itself coherent. But the theoretical model of the mountain that is realized in each of them is very different. There is a radical theoric shift from the first to the second model.

Leopold's theoric shift to the second model is only possible because he developed a new representational system, one that includes the newly created notion of "thinking like a mountain." This new system of representation is the precondition to frame the problem as suggested in the final argument. The new perspective that becomes visible in this argument counterbalances the notion of "hunters' paradise" in the first model. Even though the wish to create such a paradise might still be acceptable as a premise, it gets constrained by the new perspective.

In all three examples theoretic model abduction plays a crucial role. The model that is abductively created in each case is visible in the argumentations that I reconstructed above. Theoretic model abduction is indeed the most basic step in approaching anything, because we have to frame the subject of our attention in some way, and we do so by means of a given system of representation, and from a certain perspective. Without developing models, I would argue, we could never make sense of what is happening around us. This means that theoretic model abduction is a necessary condition for any abductively created insight.

The central role of model abduction is the starting point for my argument that we need to distinguish at least the five forms of abduction that I specified above. If every abductively created insight presupposes model abduction, then it must be possible to create also the elements abductively that occur in these models. Since these models are visible in the arguments that we must be able to construct according to my definition of insight, and since we find in these arguments propositions referring to facts, to types, and those referring to law-like relations (in the if-then statements), at least these three types of things must be abductively inferable. In addition to these three and model abduction itself we need as a fifth form the one that allows the creation of the new representational systems that are necessary for creating theoretic models.

A questionable point in my definition of "insight" might be the condition that the final outcome of abductive reasoning must be a system of beliefs that could be represented in the form of an argument (see also Aliseda 2006, 40-41). In my examples, I used only deductively valid arguments, but other forms of argument could be used as well.

The thesis that it must be possible to represent the structure of an explanatory system of beliefs as an argument can be justified by the fact that the goal of abductive reasoning is an explanation. Whatever our definition of an explanation might be, there is no question that any explanation must fulfill at least two conditions: First, it relates an explanans—that is a set of assumptions that is doing the explaining—to an explanandum, that is what needs to be explained; and, second, accepting the truth of the explanans is sufficient, for a certain person, to stop searching for further reasons to accept the explanandum. (Note that I do not claim that an explanation has to meet any objective criteria. Given the fact that even the standards for mathematical proofs changed over time (Hanna and Jahnke 1996), those criteria would probably be much harder to defend than a relativist account of explanation. Of course, there might be people who are gullible enough to accept anything as an explanation. But for my purposes it is sufficient to allow relativism and to leave room for criticizing and improving given explanations by the scientific community.)

Since accepting an explanans as sufficient for accepting an explanandum is equivalent to accepting a set of reasons as sufficient for accepting a conclusion, it is clear that claiming something to be an explanation implies that it must be possible to represent it as an argument whose premises are acceptable and whose conclusion describes the phenomenon that is in need of an explanation.

If any explanation can be represented in form of an argument, then the problem of finding an "explanatory hypothesis" in abduction can be reduced to the problem of finding that particular set of acceptable propositions which can be arranged as an argument whose conclusion is the phenomenon to be explained. Having found such an argument defines an insight that tells us that the search for explanation can stop.

"Diagram" and "Diagrammatic Reasoning"

After providing thus a solution to the first problem of abductive reasoning, we can turn now to the second one: the question of how the process of creating an explanatory hypothesis can be facilitated by means of diagrammatic reasoning. As already said, my thesis is that diagrammatic reasoning can play a crucial role as a "scaffold" for abductive creativity, if it is defined in the specific way suggested by Peirce who stressed the normative power of representational systems.

Peirce defined *diagrammatic reasoning* as a process consisting of five steps (in a letter in which he describes his discovery):

> The first things I found out were that all mathematical reasoning is diagrammatic and that all necessary reasoning is mathematical reasoning, no matter how simple it may be. By diagrammatic reasoning, I mean reasoning which constructs a diagram according to a precept expressed in general terms, performs experiments upon this diagram, notes their results, assures itself that similar experiments performed upon any diagram constructed according to the same precept would have the same results, and expresses this in general terms. This was a discovery of no little importance, showing, as it does, that all knowledge without exception comes from observation. (Peirce NEM IV 47-48)

As already mentioned, Peirce's definition of a *diagram* differs from our usual understanding that contrasts diagrams as "graphical," "pictorial," or "spatial" representations with "sentential representations." Such a distinction does not exist for Peirce. According to Peirce, even sentences and equations are diagrams. Only two features are essential for "diagrams": that they represent relations (Peirce EP II 13, 17) and that they are "carried out upon a perfectly consistent system of representation, founded upon a simple and easily intelligible basic idea" (CP 4.418).

A *system of representation* or a "diagrammatical syntax" (Peirce NEM III 162) can be defined by three elements: first, an *ontology* that defines the entities (including relations) which can be represented by means of the system; second, *conventions* that prescribe how to construct a diagram and how to read it; and third, *rules* that determine, first, which experiments with diagrams are possible and, more importantly, the necessary outcome of those experiments. An example for such a representation system is a system of axioms in mathematics like the one defined by Euclid.

If we insert this definition of a diagram into the definition of *diagrammatic reasoning* quoted above, we can determine the five steps that characterize the latter as follows:

1. Construct a diagram by means of a consistent system of representation
2. Perform experiments upon this diagram according to the rules of the chosen system of representation
3. Note the results of those experiments
4. Check the generality of these results
5. Express these results "in general terms"

The *normative* role of the chosen system of representation in the five-step process of diagrammatic reasoning becomes visible in the fact that its rules and conventions are norms that determine how to construct, read, and transform diagrams. The rules and conventions determine what is permissible in diagrammatic reasoning and what is not (Peirce NEM IV 318).

The essential point of the Peircean definition of "diagram" is that it gives up the term's etymological connection to "drawing" that it inherits from the Greek "gramma." Even though everybody seems to be happy with an understanding of "diagram" that defines the term simply in contrast to sentential representations, I don't think that this is sufficient. If a concept is supposed to fulfill a function in a theory, it must be clear which requirements are to be met for something to be counted under this concept. Peirce's criteria provide a clear answer to this challenge: A diagram is everything that represents relations and that is constructed by means of a consistent system of representation.

I think that it is an advantage of this definition that it neither draws a line between sentential and graphical, nor between external or internal representations. I cannot see how either of these distinctions could be significant when it comes to explaining abductive creativity. For the same reasons—precision and unification—I would use the Peircean notions of diagram and diagrammatic reasoning as alternatives to "model" and "model-based reasoning." I am using the term "model" only in "theoretic model abduction" where it describes the process of creating an explanatory system of beliefs that can be represented in form of an argument.

Representational Systems as a Scaffold for Abductive Insights

Many studies in cognitive science stressed as a crucial feature of "diagrammatic reasoning" that working with graphical representations constrains and guides cognitive processes (Scaife and Rogers 1996, 189; Suthers 2003; Nersessian 2008, 163). But, as Scaife and Rogers asked almost 15 years ago: How exactly "do graphical representations work" when it comes to such guidance?

Nancy Nersessian answered this question recently in a concise statement that summarizes both the literature on the cognitive functions of diagrams—understood in the traditional sense of external, graphical representations—and her own extensive case studies:

> [A] wide range of empirical data support the view that in making explicit, highlighting, or supplying structural and behavioral information, diagrammatic representations provide constraints and affordances for inferences in reasoning processes. (Nersessian 2008, 161)

While I agree that the main function of using diagrams is both to limit the search space in which someone looks for explanatory hypotheses ("constraints") and to scaffold abductive creativity by stimulating and directing the imagination ("affordances"), I would argue that it is not only the "information" that is represented in diagrams that makes both possible. The points I would stress additionally can be specified by three theses:

1. It does not matter much whether we are trying to solve a problem by means of graphical or sentential representations; the only thing that really matters is whether these representations employ the normative power of a well-defined system of representation.
2. Diagrammatic reasoning is crucial because it visualizes—whether in graphical or in sentential form—the relation between a possible explans and an explanandum.
3. Diagrammatic reasoning is crucial because it guides the experimentation with possible relations between explanans and explanandum.

In order to develop my arguments for these three theses it might be best to start with the second one. Reflecting again on our understanding of "explanation," we should not only note that an explanation can be defined as a relation between explanans and explanandum, but also that this relation is never directly observable. Even though the explanans might be observable—let's say the person that allegedly produced some footprints on the beach that we try to explain—the explanation itself is not, because an explanation refers always to the relation between explanans and explanandum, and a relation is never observable, as already Hume argued; only the relata might be observable.

In so far as we are creating something new in abduction—either a proposition describing a fact, a type, a law, a model, or a new system of representation—and since the explanation we are looking for is not directly observable, it should be impossible to perform creative abduction without having some sort of representation of both the phenomenon to be explained and of the relation between a possible explanatory hypothesis and this phenomenon. We need to handle a variety of representations to perform creative abduction, and the first function of the diagrams we are using in diagrammatic reasoning is to support this activity.

But diagrams do not only provide representations of what is otherwise unobservable, they also allow us to experiment with the unobservable. And experimentation is necessary for creative abduction, because without it we would not be able to relate a newly created hypothesis to the phenomenon we want to explain; we need to experiment with possible hypotheses and possible representations of the problem that we try to solve as long as it takes to "fit" the former into the latter. However—and this is now crucial for understanding the normative force of representation systems—we would never be able to see the significance of the results of such experimentation if these results would not be determined by the rules of the chosen system of representation. If we could not be sure that there is a necessary relation between the fact that the inner cube has six sides and the claim that at least six cuts are necessary, the theoric shift to the inner cube would mean nothing for the problem at hand. Similarly, if we would not be sure that Euclid's axioms guarantee that the angles α and α', and β and β', are equal in Figure 4, then the introduction of the auxiliary line would not have any significance whatsoever.

It is, thus, the normativity of representation systems which is crucial for the creativity that is possible in diagrammatic reasoning—not in a direct way, because the rules of the system do not create any explanatory hypothesis, but indirectly: Without knowing that experimenting with diagrams leads to necessary implications of what is represented in these diagrams, we would never be able to see the significance of what we are creating in abduction. The rules of the system of representation we are using for the construction of and experimentation with diagrams provide thus an indispensable "scaffold" for creativity. They alone can guarantee that we actually got an explanation—and not simply an unrelated proposition—when we come up with an explanatory hypothesis like "If the inner cube has six sides, then at least six cuts are necessary," or "If the dotted line in Figure 4 is parallel to the triangle's base, then $\alpha = \alpha'$ and $\beta = \beta'$."

Conclusion

The central point of my argument was that the main reason for the often observed fact that diagrams "scaffold" the creation of abductive insights can be found in the normative power of the systems of representation that we need to construct diagrams and to experiment with them. This normativity of the representation system has two effects. First, when we know from the very beginning that a successful explanation will be an explanation that can be characterized as an inference presented by the means of this system, then the problem of creative abduction can be reduced to the problem of how to come up with propositions that fit into such an inference. Second, the normativity of the representation system we are using guarantees that any experimental transformation of a diagram will lead to necessary results. This means that any experimentation forms a kind of bridge to a possible explanation: because the outcome is determined by the rules of the system, experiments prepare and prefigure already the inferential form that a final explanation will have.

References

Aliseda, A. 2006. *Abductive Reasoning. Logical investigations into discovery and explanation*. Dordrecht: Springer.

Barwise, J. and Etchemendy, J. 1994. *Hyperproof*. CSLI Lecture Notes No. 42 Stanford, Calif.: Center for the Study of Language and Information.

Bowers, K. S., Farvolden, P. and Mermigis, L. 1995. Intuitive Antecedents of Insight. In S. M. Smith, T. B. Ward and R. A. Finke, (eds.), *The Creative Cognition Approach*, 27-51. Cambridge, Mass.: MIT Press.

Cheng, P. C.-H. and Simon, H. A. 1995. Scientific Discovery and Creative Reasoning. In S. M. Smith, T. B. Ward and R. A. Finke, (eds.), *The Creative Cognition Approach*, 205-228. Cambridge, Mass.: MIT Press.

Clement, J. 2008. *Creative model construction in scientists and students : the role of imagery, analogy, and mental stimulation*. Dordrecht; London: Springer.

Dominowski, R. L. 1995. Productive Problem Solving. In S. M. Smith, T. B. Ward and R. A. Finke, (eds.), *The Creative Cognition Approach*, 73-95. Cambridge, Mass.: MIT Press.

Hanna, G. and Jahnke, H. N. 1996. Proof and Proving. In A. J. Bishop, (ed.), *International Handbook of Mathematics Education*, 877–908. Dordrecht Kluwer Academic.

Hegarty, M. 2004. Diagrams in the mind and in the world: Relations between internal and external visualizations. In A. Blackwell, K. Marriott and A. Shimojima, (eds.), *Diagrammatic Representation and Inference*, 1-13. Berlin: Springer-Verlag.

Hoffmann, M. H. G. 2005. Signs as Means for Discoveries. Peirce and His Concepts of 'Diagrammatic Reasoning,' 'Theorematic Deduction,' 'Hypostatic Abstraction,' and 'Theoric Transformation'. In M. H. G. Hoffmann, J. Lenhard and F. Seeger, (eds.), *Activity and Sign – Grounding Mathematics Education*, 45-56. New York: Springer.

Jacobs, H. R. 1970. *Mathematics. A Human Endeavor.* San Francisco: Freeman.

Josephson, J. R. and Josephson, S. G. eds. 1994. *Abductive Inference. Computation, Philosophy, Technology*. New York: Cambridge UP.

Kaplan, C. A. and Simon, H. A. 1990, In Search of Insight. *Cognitive Psychology* 22: 374-419.

Larkin, J. H. and Simon, H. A. 1987, Why a Diagram Is (Sometimes) Worth 10000 Words. *Cognitive Science* 11: 65-99.

Leopold, A. 1949. *A Sand County almanac, and Sketches here and there*. New York: Oxford Univ. Press.

Magnani, L. 2009. *Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*. Cognitive Systems Monographs, Vol. 3 Berlin: Springer.

Nersessian, N. J. 2008. *Creating scientific concepts*. Cambridge, Mass.: MIT Press.

Norton, B. G. 2005. *Sustainability. A Philosophy of Adaptive Ecosystem Management*. Chicago University of Chicago Press.

Peirce CP. *Collected Papers of Charles Sanders Peirce*. Volumes I-VI, ed. by Charles Hartshorne and Paul Weiss, 1931–1935, Volumes VII-VIII, ed. by Arthur W. Burks, 1958; quotations according to volume and paragraph Cambridge, Mass.: Harvard UP.

Peirce EP. *The Essential Peirce. Selected Philosophical Writings. Vol. 1 (1867–1893), Vol. 2 (1893–1913)*. ed. by Nathan Houser and Christian Kloesel, Peirce Edition Project Bloomington and Indianapolis 1992 +1998: Indiana University Press.

Peirce NEM. *The New Elements of Mathematics by Charles S. Peirce*. (C. Eisele, ed.), I-IV The Hague-Paris/Atlantic Highlands, N.J., 1976: Mouton/Humanities Press.

Scaife, M. and Rogers, Y. 1996, External cognition: how do graphical representations work? *International Journal of Human-Computer Studies* 45: 185-213.

Schurz, G. 2008, Patterns of abduction. *Synthese* 164: 201-234.

Smith, S. M. 1995. Fixation, Incubation, and Insight in Memory and Creative Thinking. In S. M. Smith, T. B. Ward and R. A. Finke, (eds.), *The Creative Cognition Approach*, 135-156. Cambridge, Mass.: MIT Press.

Suthers, D. D. 2003. Representational Guidance for Collaborative Inquiry. In J. Andriessen, M. Baker and D. D. Suthers, (eds.), *Arguing to learn. Confronting cognitions in computer-supported collaborative learning environments*, 27-46. Dordrecht: Kluwer Academic Publishers.