Logique & Analyse **217** (2012), 23–46

A FORMAL APPROACH TO LAKATOSIAN HEURISTICS

CAN BAŞKENT

Abstract

In this work, we offer a formal approach to analyze Lakatosian heuristics. We first give overviews of Lakatosian heuristics based on Lakatos's seminal work *Proofs and Refutations* and subset space logic which is a bimodal epistemic logic. Then, we establish the connection between Lakatosian heuristics and subset space logic by making use of an extended version of subset space logic which was suggested earlier. We then conclude with discussing the merits of our approach.

1. *Introduction*

Lakatos's *Proofs and Refutations* (PR, henceforth) exhibits a careful analysis of a very significant mathematical development, namely the evolution of Euler's Theorem $V - E + F = 2$ for three dimensional polyhedra where V, E and F are the number of vertices, edges, and faces of the polyhedron respectively.

In his canonical piece, Lakatos presented a genuine contribution to philosophy of mathematics with his deep focus on the actual history of the theorem starting from Cauchy's well-known proof. In his analysis, Lakatos introduced several notions in order to be able to give a rational account of the historical and the methodological development of the theorem throughout its historical course. In this regard, for Lakatos, the *development* of a mathematical theorem together with its proof was a very significant aspect of the growth of knowledge in mathematics. As Kiss put it, "in Lakatos's heuristics, the theorem is not ready when we start to prove it. It is stated in a possibly false generality, and it can be formulated several times in the process [of its development]." [4]. In other words, in Lakatosian heuristics, one starts off with rather loose and overly generalized statements and makes them more and more precise *along the course of their development*. This is one of the reasons which makes the discussions on Lakatosian heuristics indispensable to philosophy of logic and mathematics.

In this work, our aim is to utilize a rather weak but expressive epistemic logic with geometric semantics to formalize the Lakatosian heuristics. First, we recall Lakatos's methodology of mathematics based on his acclaimed work *Proofs and Refutations*. Second, we briefly introduce an epistemic logic which we will use along these lines. After providing the necessary technical background, we then establish the connection between the two. Then, we discuss the philosophical significance of our approach. Finally, we will conclude with some remarks and point out several research directions for future work.

What we achieve in this paper is essentially as follows. By utilizing a formal system to analyze Lakatosian heuristics, we aim at to point out the computational aspects of his approach: a step by step procedure describing how Lakatosian methodology works. This observation shall have an unexpected outcome within the domain of Lakatosian philosophy. Let us now start with reviewing Lakatos's methodology.

2. *Lakatosian Heuristics*

2.1. *Basics of Lakatosian Heuristics*

Lakatosian methodology follows a simple yet well-defined road map which consists of the following methodological steps [3].

(1) Primitive conjecture.
(2) Proof (a rough thought experiment or argument, decomposing the primitive conjecture into subconjectures and lemmas).
(3) Global counterexamples.
(4) Proof re-examined. The guilty lemma is spotted. The guilty lemma may have previously remained hidden or may have been misidentified.
(5) Proofs of the other theorems are examined to see if the newly found lemma occurs in them.
(6) Hitherto accepted consequences of the original and now refuted conjecture are checked.
(7) Counterexamples are turned into new examples, and new fields of inquiry open up.

Let us now elaborate further on these steps. Lakatos employed three main strategies to implement the method of proofs and refutations: *monster-barring*, *exception-barring* and *lemma incorporation*. A lengthy quote we give in the Appendix illustrates their use explicitly. Nevertheless, before giving a formal account of the aforementioned strategies, let us first explicate them further.

The method of "monster-barring" deals with the objects which are *not in mind* when the conjecture is put forward. They are, in this sense, monsters and should be excluded from our domain of discourse. In terms of Euler's formula, we can think of them as the objects "of no theoretical interest, [since] no normal mathematician would ever think of them as polyhedra." [7]. The second strategy, namely the "exception-barring" accepts that the theorem in its stated form is not valid due to the emergence of some genuine counterexamples targeting the correctness of the theorem itself. In other words, the initial domain of the objects which was previously thought of satisfying the Euler's theorem is discovered to be too large. Thus, these counterexamples should be excluded from the original domain which results in a contraction of the domain. The contraction of the domain, we need to underline, is necessitated by the emergence of the genuine counterexamples. The third and the last method is called "lemma incorporation". Lemma incorporation describes the last item in the above list. To put it in a different way, lemma incorporation depicts the way we turn the counterexamples into new examples, and those new examples are helpful for the modified and reformulated version of the theorem. This procedural description of theorem formation is significant for many other purposes as well. Namely, such a procedural methodology can be given an algorithmic method [12].

2.2. Proofs and Refutations

In this work, we focus on *Proofs and Refutations* which was first published in the *British Journal of Philosophy of Science* in 1963 and 1964 as four parts, and then appeared as a book in 1976.

What made PR an easy read and a distinguished work of philosophy is perhaps its presentation in dialogue form. The dialogue took place in a rather advanced classroom setting where the students discussed the Euler Conjecture with some facilitation from their teacher, and consequently came up with several proofs and refutations; hence the name of the essay. What made this class an advanced one is the fact that the conjectures, counterexamples, proofs and refutations which were put forward by the students had been taken from the actual history of the conjecture.

In PR, one can easily observe that counterexamples play a very significant role. They inspire proofs or disproofs, or sometimes a new formulation of the conjecture. However, Lakatos presented the counterexamples in a dialectic and a didactic fashion without mentioning their heuristic roles. A counterexample, for Lakatos, can be of either positive heuristics or negative heuristics. The difference between these two separate roles of counterexamples can be summarized as follows. "Positive heuristic contains rules that help us in application, in the formation of a fallible version of the hard core [of the theorem]. (...) The role of negative heuristic is simply to defend the hard core maybe in a less creative way." [5]. A detailed account of all the counterexamples mentioned in PR was already given in a previous work [2]. Thus, we refer the interested reader to the aforementioned article.

It is worthwhile to note that Lakatos's presentation of the mathematical arguments in a dialog form immediately implies a possible use of a game theoretical approach for further analysis of the concepts in question. One can also recognize that the notion of *strategy* is embedded in the Lakatosian heuristics. The order in which the arguments were put forward, and the way that these arguments were treated define the rules of the game, and the aforementioned three major heuristic methods can then be considered as game theoretic strategies.

Nevertheless, there is a critical point. It has been claimed that Lakatos's *rationally reconstructed account* of the history of the development of the Euler theorem often diverged from the actual history of the subject. Koetsier stated that "there is no doubt that *Proofs and Refutations* contains a highly counterfactual rational reconstruction." [6]. In order to be able to stick with our current agenda here, we will not go into much of such historical details, and therefore, refer the interested reader to the aforementioned reference for further expositions. Thus, our focus here is restricted to Lakatos's heuristics as it was presented in PR.

Now, we can discuss the major aspects of PR. The main conjecture which was discussed in PR is the following.

$$V - E + F = 2$$

for all polyhedra, where V, E and F denote the number of the vertices, the edges and the faces of the given polyhedron respectively. This conjecture is often called the *Descartes–Euler conjecture* for historical reasons. However, following the original work, we will refer to it simply as *Euler's Theorem*. In this context, the integer value which is obtained from the equation $V-E+F$ for some polyhedron P is called the *Euler characteristic* of P.

The proof of this theorem, as it was stated in PR, is due to Cauchy. Let us summarize it step by step.

Step 1 Imagine that the polyhedron is hollow and made of a rubber sheet. Cut out one of the faces and, strech the remaining faces to a flat surface (or board) without tearing. In this process, V and E will not alter. However, as we removed a face, the Euler characteristics of the polyhedron has decreased by one. Therefore, we now need to show that $V - E + F = 1$.

Step 2 Triangulate the remaining map. Drawing diagonals for those curvilinear polygons will not alter $V - E + F$ since E and F increase simultaneously by the same amount while V does not change.

Step 3 Remove the triangles. It can be done in two ways: either one edge and one face are removed simultaneously; or one face, one vertex and two edges are removed simultaneously. During this process, $V - E + F$ remains unchanged. Consequently, at the end of this process, we will end up with an ordinary triangle for which $V - E + F = 1$ holds trivially.

Observe that there are three lemmas that have been used implicitly throughout the proof. Focusing more on these lemmas will help us analyze the proof better.

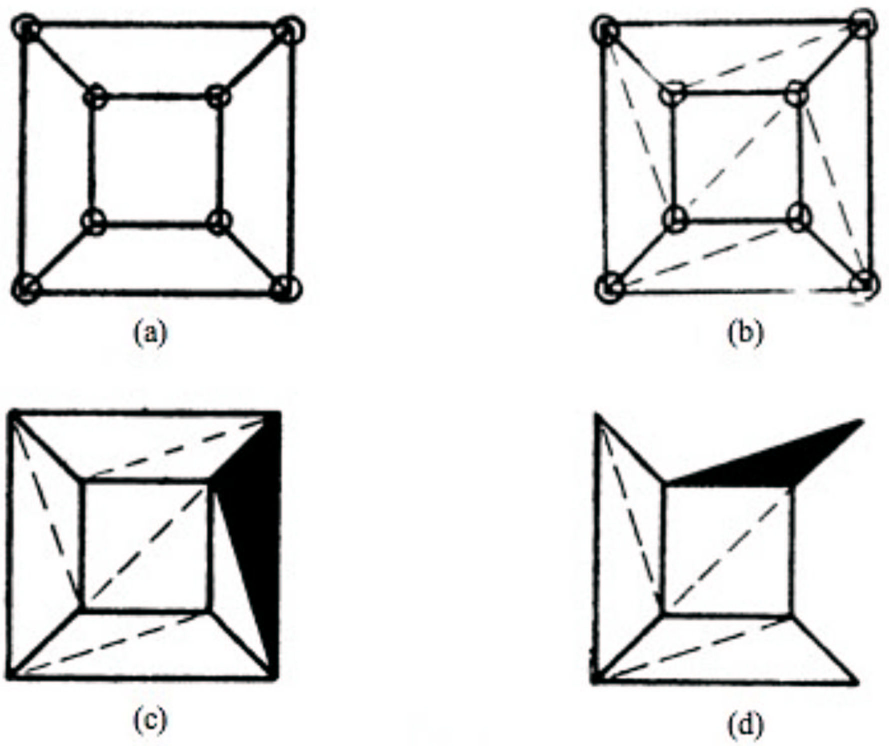

Figure 1. Graphical depiction of Cauchy's proof taken from [8].

Lemma 1: *Any polyhedron, after a face is removed, can be stretched flat on a flat surface.*

Lemma 2: *In triangulating the map, one will always get a new face for every new edge.*

Lemma 3: *There are only two alternatives for removing a triangle out of the triangulating map: the disappearance of one edge or else of two edges and a vertex — when one decreases the number of triangles by one. Furthermore, one will end up with a single triangle at the end of this process.*

The very first counterexamples emerge upon the three lemmas used in the proof. *Local counterexamples* deny the specific lemmas or the constructions which were used in the proof without targeting the main conjecture itself. However, *global counterexamples* deny the main conjecture, and without even consulting the proof, demonstrate that the conjecture is false. A pair of nested cubes, urchin, cylinder and picture frame are some of the global counterexamples mentioned in PR; and in this work, we focus on global counterexamples as they directly target the conjecture.

As we have discussed above, the method of monster-barring is supposed to deal with such cases. At this stage, Lakatos discussed an intuitive but quite straight-forward account of geometric topology. For instance, the picture-frame (i.e. torus) cannot be inflated into a sphere or cannot be stretched onto a plane. The reason for that is the fact that the *genus* of the sphere is zero (i.e. it has no holes), even if we remove a single point from the sphere. Removing a single point from a sphere, on the other hand, makes it possible to stretch it onto the Euclidean plane. Moreover, the sphere with one single point removed is homeomorphic to the Euclidean plane — there exists a bicontinuous isomorphism between the two. Thus, it is topologically the "same" to stretch a polyhedron onto the Euclidean plane or onto a sphere with one point removed. Then, it was concluded that the picture-frame could be inflated only onto a torus, not onto a sphere. It should now be recalled that the torus has genus one, namely it has one hole passing through it. Indeed, the general formula of the Euler characteristics in (oriented) manifolds is the following.

$$V - E + F = 2 - 2.g(S)$$

where S is the surface we consider on which the polyhedron will be inflated, and $g(S)$ is the genus of the surface S.

This equation is the exact reason why the picture frame is a global counterexample to Euler's conjecture in its stated form. Yet, since the torus is not simply connected (In other words, on torus, not every simple and closed curve can be continuously shrunk to a point. Consider a closed path which encloses the "hole".), then it was argued that the original Euler conjecture holds only for simple polyhedra, namely for those when having a face removed, could be stretched onto a plane. In this way, the domains of the conjecture and Lemma 1 are restricted. However, the proof still remains the same.

The discussion we have presented so far exhibits the central idea of Lakatos's heuristics. We start with a proposition (Euler's Conjecture), and then restrict its domain to those for which the conjecture holds. Then, by monster-adjustment and exception-barring, we reconstruct our epistemic model. The newly reconstructed model, in this case, depends on the previous one.

The existence of the geometric objects whose Euler characteristics are not 2 suggests that the use of *possible worlds* to represent Lakatosian heuristics is worth pursuing. In other words, we can very well imagine that there can be a mathematical and ontological state at which the Euler characteristics of a given object is not equal to 2. Therefore, modal models with possible worlds can be used to formalize Lakatosian methodology with the aforementioned idea in mind.

Moreover, the dynamic aspect of updates and the theory changes encourage us to use an epistemic modal logic. However, instead of utilizing the well-known Kripkean semantics, we will use an epistemic modal logic with geometrical semantics for our purpose.

3. *Subset Space Logic*

3.1. *Basics of Subset Space Logic*

In order to give a formal account of the Lakatosian heuristics, we use subset space logic which was first put forward in early 90s by Moss and Parikh. Their goal was to present a bimodal logic called *subset space logic* (SSL, henceforth) to formalize reasoning about sets and points [11]. The language of SSL has two modal operators where one of them is intended to quantify *over* the sets ($\Box$) whereas the other *in* the sets (K). The subsets in Moss and Parikh's structure are called *observation* or *measurement* sets, and the underlying motivation to introduce of these two modalities is to be able to speak about the notion of *closeness* — which is a concept familiar from topological reasoning. The key idea of Moss and Parikh's approach to the concept of closeness can be formulated as follows.

> In order to *get close*, one needs to make some *effort*.

This idea establishes the connection between knowledge and effort in SSL. Therefore, to gain *knowledge*, we need to make some *effort*. By spending some effort, we eliminate some of the existing possibilities, and obtain a smaller set of possibilities. The smaller the set of observations is, the larger the information we have. However, note that there is no given explicit way for *making the observations finer*. We will elaborate more on this later.

Let us now specify the technical details of SSL. The language of subset space logic $\mathcal{L}_{SSL}$ has a countable set P of proposition letters, a truth constant $\top$, the usual Boolean operators $\neg$ and $\wedge$, and two modal operators K and $\Box$. The formulas in $\mathcal{L}_{SSL}$ are obtained from atomic propositions by closing them under $\neg$, $\wedge$, K and $\Box$. A *subset frame* is a pair $\mathcal{S} = \langle S, \sigma \rangle$ where S is a non-empty set of points, and σ is a set of subsets of S, i.e. $\sigma \subseteq \wp(S)$. However, note that σ is not necessarily a topology. The elements of σ are called *observations*. The triple $\mathcal{S} = \langle S, \sigma, v \rangle$ is called a subset space model where $\langle S, \sigma \rangle$ is a subset frame, $v : P \to \wp(S)$ is a valuation function. We can now define the semantics of subset spaces.

Definition 3.1: *For $s \in S$ and $s \in U \in \sigma$ in the subset space model $\mathcal{S} = \langle S, \sigma, v \rangle$, we define the satisfaction relation $\models_{\mathcal{S}}$ on $(S \times \sigma) \times \mathcal{L}_{SSL}$ by induction on the length of the formulas. We will drop the subscript $_{\mathcal{S}}$ when the subset space model we are in is obvious.*

$$
\begin{array}{lll}
s, U \models p & \textit{iff} & s \in v(p) \\
s, U \models \varphi \wedge \psi & \textit{iff} & s, U \models \varphi \quad \textit{and } s, U \models \psi \\
s, U \models \neg\varphi & \textit{iff} & s, U \not\models \varphi \\
s, U \models \mathsf{K}\varphi & \textit{iff} & t, U \models \varphi \quad \textit{for all } t \in U \\
s, U \models \Box\varphi & \textit{iff} & s, V \models \varphi \quad \textit{for all } V \in \sigma \textit{ such that } s \in V \subseteq U
\end{array}
$$

We call $\Box$ the *shrinking* operator and K the *knowledge* operator. The duals of $\Box$ and K are $\Diamond$ and L respectively, and defined in the usual way: $\mathsf{L}\varphi \equiv \neg\mathsf{K}\neg\varphi$ and $\Box\varphi \equiv \neg\Diamond\neg\varphi$.

The pair (s, U) is called a *neighborhood situation* if U is a neighborhood of s, i.e. if $s \in U \in \sigma$. If at (s, U) we *know* φ, this then means that we can move from the given reference point s to any other point t in the given neighborhood U without changing the neighborhood. Likewise, by using the shrinking modality, we shrink the neighborhood U around the given point s to another subset V to obtain a new neighborhood situation (s, V) keeping the reference point s unchanged.

The reason as to why we use SSL to express Lakatosian heuristics is due to its bimodal structure. The epistemic modality helps us to express the knowledge, and the effort modality enables us to formalize how we obtain the knowledge or how we improve the epistemics of the agents. Therefore, SSL connects how knowledge and effort interact. Moreover, we can use the effort modality to express a variety of different notions: observations, measurements, calculations, computations, experiments etc. — namely the procedures that can affect the knowledge of the knower [1, 11].

The axiomatization of the SSL simply reflect the fact that the K modality is $\mathsf{S5}$ whereas the $\Box$ modality is $\mathsf{S4}$. Moreover, we have two additional axioms to state the interaction between the two modalities (called cross axiom); and the independence of real world from knowers point of view (called atomic permanence) [11]. We refer the reader to the original paper for the detailed treatment of the technical results including the completeness and decidability results.

Notice that in SSL, the shrinking modality is a dynamic modality. Any improvement in the knowledge situation of the agent can be represented by the shrinking modality $\Diamond$. However, considering its semantics, we can immediately see that it also suffers from "$\exists$-sickness. Modal symbol $\Diamond$ states that there *exists* a subset, but which subset is it?

3.2. *Controlled Subset Spaces*

The problem we focus on can be thought of an instantiation of the "$\exists$-sickness" problem. Recall that the $\Diamond$ operator only states that "there exists" a subset of the given observation set, but does not precisely indicate which subset is the intended one. In an earlier work, a system in which such subsets can be specified precisely was suggested [1]. Due to the explicit nature of that logical structure, such systems were called "Controlled Subset Spaces". In order to make the present work here a bit more self-contained, let us briefly describe the controlled subset spaces here. We refer the interested reader to the original work for the motivations and problems which have led to the controlled subset spaces [1].

Let us start with defining the basic notions which we will use along these lines. The image set fU of a function f under the set $U \subseteq S$ will be the defined as $fU = \{f(x) : x \in U\}$ for $f : S \to S$. Furthermore, we call f a *contraction mapping*, if for every subset $U \subseteq S$, we have $fU \subseteq U$. Let $\mathcal{F}$ be an arbitrary collection of contraction mappings which are defined on S, and further let $F \subseteq \mathcal{F}$ be some selection of such contraction mappings. Given a subset space $\mathcal{S} = \langle S, \sigma, v\rangle$, we obtain the *image space* $\mathcal{S}_F = \langle S, \sigma_F, v\rangle$ where $\sigma_F := \{fU : f \in F, U \in \sigma\}$. In this setting, the valuation v does not change. Here, $\mathcal{F}$ can be considered as all possible contracting mappings for that particular model and space, whereas F is the set of admissible mappings chosen from $\mathcal{F}$.

As we have already underlined, each function $f \in F \subseteq \mathcal{F}$ is a contraction mapping which was intended to represent the increase in knowledge. Hence, $fU \subseteq U$ should hold for each function $f \in F$ and for each observation set $U \in \sigma$. On the other hand, observe that each $V \in \sigma_F$ is the image of an observation set $U \in \sigma$ under some function $f \in F$. Given a neighborhood situation (s, U) in $\mathcal{S}$, we will get another neighborhood situation (s, fU) in $\mathcal{S}_F$ for some $f \in \mathcal{F}$ by removing the points in U which are not in the image of U under f. But, is s in fU? Because otherwise (s, fU) would not be a neighborhood situation. Therefore, we have to *force* this condition. In other words, we can only evaluate the formulas at (s, fU) only if $s \in fU$. Nevertheless, this imposition has a natural intuition in our setting since elimination of possibilities in the current observation set does not necessarily change the actual (possible) world. Also notice that for the contracting mappings $f, g \in F$, we have $f(g(U)) \subseteq g(U) \subseteq U$ by set up. Thus, consecutive application of contracting mappings is allowed and captured in our framework. We can now give the definition of controlled subset spaces.

Definition 3.2: *$\mathcal{S} = \langle S, \sigma, v, \mathcal{F}\rangle$ is called a controlled subset space where S is a set, σ is any collection of the subsets of S, $v : P \to \wp(S)$ is a valuation function and $\mathcal{F}$ is a collection of contraction mappings $\mathcal{F} = \{f : f$ is a contracting mapping and $f : S \to S\}$ defined on S.*

We now introduce an additional modality $[F]$ representing the controlled shrinking. The intended meaning of the statement $[F]\varphi$ is that "after the application of each function $f \in F$, φ becomes true". Note that, after application of the function f, we evaluate the formula φ in the new space $\mathcal{S}_F$. The change in the spaces is essential to reflect the underlying idea of controlled shrinking. On the other hand, the dual of $[F]$ is denoted by $\langle F\rangle$ and defined in the usual way. The semantics of the controlled subset spaces is not different from the original subset spaces except from the controlled shrinking modality which is evaluated in the image space $\mathcal{S}_F$ of $\mathcal{S}$:

$$s, U \models_{\mathcal{S}} [F]\varphi \quad \text{iff} \quad s, fU \models_{\mathcal{S}_F} \varphi \text{ for each } f \in F$$

The dual of $[F]$ will be defined as follows following the usual duality.

$$s, U \models_{\mathcal{S}} \langle F\rangle\varphi \quad \text{iff} \quad s, fU \models_{\mathcal{S}_F} \varphi \text{ for some } f \in F.$$

For simplicity, we can consider the image space $\mathcal{S}_F$ a subspace of the given space $\mathcal{S}$, and evaluate the formulas in $\mathcal{S}$. This simplification does not make any difference in the semantical evaluation of the formulas.

In the aforementioned work, some properties of controlled subset spaces were investigated, thus we refer the reader to that work for further expositions and technicalities [1].

4. *A Formalization of Lakatosian Heuristics*

4.1. *A Proposal*

Based on the above observations, we now present a formal model for Lakatosian heuristics. We consider Euler's conjecture for three dimensional polyhedra, and use controlled subset spaces to give a formal account of it.

Let us start with the method of monster barring. Recall that monster barring restricts the domain of the objects which were initially supposed to satisfy the given conjecture. However, the "contraction" of the domain is not random as, for Lakatos, it is governed by an observation, an idea, a computation or a thought-experiment, or some combination of them. In order to illustrate how controlled subset spaces help us to express Lakatosian ideas, let us start with the constructions mentioned in the lengthy quote which we reproduced in the appendix.

Consider the observation set $U = \{s, n, u, c, p, k, t\}$ where s, n, u, c, p, k, t represent the sphere, the nested-cubes, the urchin, the cylinder, the picture-frame, the cube and the twin-tetrahedra respectively. The set U is the set of possible objects which can be seen as possible worlds at which some formulas about them are valid or not. Let us assume, for simplicity, that the current state that we are occupying is s; thus U is the observation set for the agent at the state s. The agent can very well consider, and even further imagine some other objects in his observation set and thus reason whether those objects would satisfy the Euler conjecture in its given form. In this situation, the agent can increase his knowledge by discarding some possible worlds. In other words, the way the agent can improve his knowledge is to shrink U. But how can we achieve this in the Lakatosian heuristics?

Let f be the function which returns the input object x as the output only if the given object x satisfies the Euler conjecture $V(x) - E(x) + F(x) = 2$. More precisely, f is given as follows.

$$f(x) = \begin{cases} x & : \text{ if } V(x) - E(x) + F(x) = 2 \\ \text{undefined} & : \text{ otherwise} \end{cases} \tag{1}$$

Observe that f is a well-defined contraction mapping. The underlying motivation to define f as such is to mimic the characteristic function of the set of objects whose Euler characteristics are 2. This observation is perfectly consistent with the method of monster barring since the objects whose Euler characteristics are not 2 are considered as monsters and thus, should be excluded from our domain of inquiry. We achieve this by a use of a function.

Recall that for the sphere, we have $f(s) = 2$, for the nested cube $f(n) = 4$, for the urchin $f(u) = -6$, for the cylinder $f(c) = 1$, for the picture-frame $f(p) = 0$, for the cube $f(k) = 2$, and finally for the twin-tetrahedra $f(t) = 3$. Let us now construct $f(U)$, namely the image set of the present observation set U under the function f. After a brief observation, we see that $f(U) = \{s, k\}$ as only s and k have the Euler characteristics 2. Therefore, starting off from the given set of polyhedra, we *filter* the set in such a way that only the objects with the Euler characteristics 2 will remain. Therefore, we formalize this situation as follows $s, U \models [f]\chi$ where χ the given conjecture — namely $V(x) - E(x) + F(x) = 2$. As we observed, χ is true at sphere and cube.

If we want to consider some other Euler characteristics, we only need to include them as functions. In a similar fashion, let f' be the contraction mapping for the Euler characteristics 0. Similarly, the precise definition is as follows.

$$f'(x) = \begin{cases} x & : \text{ if } V(x) - E(x) + F(x) = 0 \\ \text{undefined} & : \text{ otherwise} \end{cases} \quad (2)$$

In a similar fashion, here, f' returns only the objects whose Euler characteristics is 0. We can therefore write $s, U \models \langle F \rangle \chi$ where $F = \{f, f'\}$, and χ is as before. Moreover, we can also utter $s, U \models \langle F \rangle \chi'$ where χ' is the conjecture $V(x) - E(x) + F(x) = 0$. Examples can be multiplied in a similar manner.

There are several reasons why the above constructions fit well into the context of subset space logic. The very first observation is the fact that the contraction requires some *effort*, namely the computation and the verification of the equation $V(x) - E(x) + F(x) = 2$. Subset space logic, in its original presentation, does not include any operator which precisely designates the behavior of the contraction. From Lakatosian point of view, the construction is not yet over. We also spend additional effort to come up with the *rule* of the function. This is what Lakatos usually calls a *thought experiment*. We formalize this additional effort by constructing the set F of functions.

The discussion of the method of monster barring in PR focused on the formula χ as it was the conjecture that they initially set out to prove. As we demonstrated, this method would have been perfectly applicable to the present case even though our initial conjecture were different (say χ').

The crucial point of referring to some polyhedra as "monsters" is also captured in this framework. Recall that the monsters can be turned into the examples of the modified conjecture. In our previous illustration, for instance, t is a monster. Yet, it can be an example of the conjecture χ' (which is controlled by f'). Therefore, having a large enough collection

F of mappings, we can represent the method of monster barring in a sound fashion.

The second strategy of the methods of proofs and refutations is the method of exception barring. Exception barring works exactly the same as the method of monster barring up until the point of turning monsters into the examples of some other conjecture. Exception barring, in this sense, stops after the domain is contracted.

The third strategy is perhaps the most crucial one. The method of lemma incorporation suggests us to extend our set of functions in consideration. This is precisely what we did for torus in the above illustration. We incorporated the lemma in such a way that f' will work for t. One of the ways to achieve this to introduce the conditions that stems from the modified lemma into the formulation of the function. For instance, in PR, the notion of genus was introduced to discuss non-simple polyhedra. Then the general form of the Euler conjecture, as we discussed previously, becomes $V(x) - E(x) + F(x) = 2 - 2.g(x)$ in the oriented objects such as torus where $g(x)$ is the genus (i.e. the number of holes) of the object x. Based on this reformulation, we can restate the function f given in the Equation 1 as follows.

$$h(x) = \begin{cases} x & : \text{ if } V(x) - E(x) + F(x) = 2 \wedge g(x) = 0 \\ \text{undefined} & : \text{ otherwise} \end{cases} \tag{3}$$

In a similar fashion, we can incorporate the lemma which addresses the simplicity condition to the Equation 2. The following equation does the job.

$$h'(x) = \begin{cases} x & : \text{ if } V(x) - E(x) + F(x) = 0 \wedge g(x) = 1 \\ \text{undefined} & : \text{ otherwise} \end{cases} \tag{4}$$

In this way, we have restricted our domain even further by incorporation more conditions in the formulation of the controlling mappings. More precisely, we now have that $F = \{f, h, f', h'\}$ in our set of applicable mappings.

As we have underlined earlier, this analysis can be incorporated into the cases where the order of the methodological steps are of importance. Namely, consecutive and ordered application of a variety of, say, counterexamples, can be captured in this framework as it is not necessarily the case that $f(g(U)) = g(f(U))$ for contracting mappings f and g. Therefore, the order of the application of contracting mappings matters, and each consecutive step towards the generation of knowledge can be expressed by a different contracting mapping in order to obtain a more complex yet explicit statement. We leave this immediate extension of our framework to the reader.

The constructions we have presented hitherto give us sufficient tools to formalize the method of proofs and refutations. We now briefly consider a single case to exemplify all of the discussions we have had so far.

4.2. *An Application*

Let us now consider torus t, and see how the Lakatosian heuristics for the Euler conjecture instantiated to torus can be formalized by using the controlled subset space logic. Based on the above illustrations and formulations, our job is now easy. We suggest that the following equation

$$t, U \models [f]\chi \vee \langle F \rangle \chi' \tag{5}$$

is sufficient to express Lakatosian heuristics in this context.

Starting off from the torus t in the observation set U, we observe and calculate that the torus does not satisfy the Euler formula $V - E + F = 2$. Thus, we need to modify the original formula f (cf. Equation 1) to get a new formula f' (cf. Equation 2) and construct the set of formula $F = \{f, f'\}$. Then, among the functions in F we can pick the correct one that is satisfied at the actual state t. In other words, among the set of possible formulas/observations, we pick the one that works at the actual state, so we pick $V - E + F = 0$.

The formalization of Lakatosian heuristic methodology by (exclusive) disjunctions gives us further hints as to how verification works within this framework. In other words, given a statement such as Equation 5, we now know to express conjecture improvement and verification. This example also illustrates how the new submodels or substructures obtained in Lakatosian heuristics and their connection with the initially given model.

4.3. *Generalization*

Based on the observations and the applications we have presented hitherto, we can now give a general account of Lakatosian heuristics for arbitrary formualae. Let $T(\vec{x})$ be the theorem in question with free variables $\vec{x}$. We can formalize the Lakatosian heuristics of the development of the theorem $T(\vec{x})$ with the input vector $\vec{x}$ as follows. For simplicity let us assume that the current epistemic neighborhood situation we are in is given (s, U).

$$\varphi(\vec{x}) = \begin{cases} \vec{x} & : \text{ if } T(\vec{x}) \text{ holds} \\ \text{undefined} & : \text{ otherwise} \end{cases} \tag{6}$$

The simple idea behind this formalization is to define knowledge emprically. In the contracted neighborhood situation $(s, \varphi U)$ we know the propositions which are emprically verified under the function φ. This reading of the controlled subset spaces agrees perfectly with the Lakatosian understanding of mathematics as a *quasi-emprical* science. Observe that the current focus in the controlled subset spaces is on the verification, not on the discovery. Among the set of functions that *may* work, we try to find and verify the one that works.

All these discussions and argumentation motivate us to set some terminology. At a given state s with a corresponding observation set U with $s \in U$ and a set F of contraction mappings $f_1, f_2, \dots \in F$, we will call f_i mappings *the epistemic methods* which are applicable to s in U. After an application of f_i to U, we will say that U *is controlled by* f_i. If U is controlled by every f_i in F, we will say U *is controlled by* F. This natural terminology is meant to reflect the fundamental ideas behind the controlled shrinking.

4.4. *Discussion*

Our logic based approach can be criticized for ignoring some issues in philosophy of language which Lakatos mentioned in PR. In other words, here, we do not argue on Lakatos's discussion of "language statics vs language dynamics" in the later parts of PR. By the same token, we do not attempt to explicate Lakatos's discussion on "terms" and "defining terms" within the context of the Euler Conjecture by using a formal language. The very first reason for us to avoid this discussion is simple: formal modal logic is far from expressing dynamic aspect of scientific discovery and verification in the sense of Lakatosian method of proofs and refutations. Meaning of the terms changes, but, in most formal systems including the one that we use, the interpretations and the semantics of the terms are conservative and do not change in time. Moreover, it must be noted that we are not *exemplifying* a mathematical discovery by using formal logic. We are *using a formal language* to clarify and explicate the Lakatosian heuristics. Therefore, our use of logic is not intended as a model for mathematical discovery that Lakatos would immediately disagree with [8, p. 143].

Then, one can very well ask: "Then, what would be the point of formalizing Lakatosian heuristics by using modal logic?" This is an important question that we cannot dispense immediately.

First of all, Lakatosian understanding of a theorem has an intensional computational flavor. As Kiss pointed out, "in Lakatos's heuristics, the theorem is not ready when we start to prove it. It is stated in a possibly false generality, and it can be formulated several times in the process [of its development]" [4]. Therefore, we can very well claim that, for Lakatos, a theorem can be thought of as a procedure, a computation or a program. This approach has been taken by computer scientists already who claim to formalize and even program Lakatosian heuristics [12]. We take one further step, and claim that a theorem, in Lakatos's sense, is indeed a program. It can very well be a program with a bug, run-time error or a mistake. It can be a program that does not terminate with an infinite loop. It can be a program that returns a wrong output. Yet, that is perfectly fine as we can have *wrong theorems* and *proofs that do not prove* in Lakatosian methodology. Therefore, if the theorem/program does not work, we may need to *fix* or *debug* the program which requires some effort, computation, test runs, trial and error process, redefinition of some concepts, variables etc. Thus, the Lakatosian meta-model of mathematical practice do indeed possess a formal structure. It is indeed a *software* in Parikh's sense [11]. What we have achieved in this work was to present a framework for that software that Lakatos reconstructed historically and rationally. In this respect, the SSL has a perfect tool, namely the box modality, to express computation and effort and their epistemic consequences. Yet, the way it computes is not explicitly encoded in the box modality, therefore the more explicit language of controlled subset spaces with its additional modality was needed. By the use of such a language, we can directly observe, manipulate and compute the procedures in the method of proofs and refutations. However, this is not the only aspect of our approach that makes it unique.

Moreover, computation oriented readings of Lakatosian method have another agenda of exposing the *un-Lakatosian* core of Lakatosian heuristics. Clearly, this claim is rather paraconsistency inclined. As Priest pointed out several times, science and mathematical practice has natural inconsistencies in it [13]. We claim that revealing the un-Lakatosian aspect of Lakatosian methodology corresponds to the paraconsistent aspect of such mathematical practice. The contradiction is there, yet, it does not prevent us from making sensible inferences. Using the terminology of paraconsistency, the un-Lakatosian meta-theory of the Lakatosian method is the contradiction that does not lead to absurdity. It should not be surprising to note that Lakatos himself pointed out several instances of inconsistencies in science including Bohr's 1913 paper as it was "inconsistently grafted on to Maxwell's theory" [9, pp. 56–8, 126][1].

Moreover, we claim that Lakatosian practice based mathematics and its heuristics can have a formal and algorithmic structure in the meta level. Lakatos makes it extremely clear that such formalism does not exist in the object level of mathematical discovery, but, he does not mention if there can be found any such structure in the meta level of his theory of methodology of scientific research programs or his method of proofs and refutations. In other

[1] We are grateful to the anonymous referee for pointing this out.

words, we claim that our computational and algorithmic approach presents the archeology of the hidden weak formalism in Lakatosian heuristics. The logical structure exposes and clarifies this hidden component. This is also directly related to our aforementioned take of Lakatosian methodology as a game theoretical set of admissible strategies. The *game of proving* can be considered as a game between prover and refuter where each makes a move by presenting an affirmative or refutative example. The game is won when the opponent has no move to make.

However, our formalism can be criticized by strict Lakatosians for being deductive. Lakatos wrote in the appendix of PR as follows.

> Euclidean methodology has developed a certain obligatory style of presentation. I shall refer to this as 'deductivist style'. This style starts with a painstakingly stated list of *axioms, lemmas* and/or *definitions*. The axioms and definitions frequently look artificial and mystifyingly complicated. One is never told how these complications arose. The list of axioms and definitions is followed by the carefully worded *theorems*. These are loaded with heavy-going conditions; it seems impossible that anyone should ever have guessed them. The theorem is followed by the *proof*. [8, p. 142 (his emphasis)]

Now, we have to be careful. We do not propose our logical system as a way to do deduction, or in general, do mathematics. As we emphasized earlier, we use its language and its formal strength to explicate Lakatos's heuristics. Therefore, our efforts do not fall into the scope of Lakatos's critique.

Let us summarize. Our approach exposes the computational, algorithmic and even game theoretical structure of the Lakatosian methodology. If the Lakatosian system has such a formal framework at the meta-level, then this formalism may make it *un-Lakatosian*. Thus, we claimed that the algorithmic and computational aspect of his methodology is the un-Lakatosian aspect of his theory.

4.5. *What Is Missing?*

Notice that our method does not give an account of how scientific knowledge is generated. Therefore, for instance, the discovery of incorporating the notion of genus to the Euler's formula cannot possibly be captured or predicted in our framework, if not in any formal frame-work. Any framework discussing such discoveries necessarily needs to exhaust all possible properties that its objects possess, and establish the connections between those objects and the set of properties. Thus, from the Lakatosian perspective, this Euclidean project of giving a formal account of mathematical *discovery* from a quasi-axiomatic perspective is not achievable. Thus, our project of approaching Lakatosian heuristics formally misses this point *per se*.

Furthermore, the notion of "concept generation" and some other linguistic issues such as "language statics vs language dynamics" that Lakatos mentioned are far from being formalized. In other words, the methodological power of choosing the admissible set of contracting mappings F out of set of mappings from the modal operator $\mathcal{F}$ or generating the set F from a mapping f is not a mathematical operation for Lakatos, but rather a context dependent, practice based empirical procedure. Therefore, our mathematical treatment cannot tell, based on the function names, which mappings should be incorporated into our methodological toolbox. This is one of the reasons that makes our approach a descriptive one. Therefore, given a theory, based on our model, one cannot decide how to apply which methodology to which object.

Nevertheless, these standard criticism towards our formal approach do not shed doubt towards the mathematical consistency of the system as SSL and its extension all are well-defined systems. Even if we switch to paraconsistent systems to expose the Lakatosian methodology as we have implied, the well-behaved contradictions still enable us to make meaningful deductions and inferences without reaching a trivial theory.

5. *Conclusion*

Applying computational methods to philosophy and methodology of science is not a new idea. An early work on the very same subject, for instance, suggested an algorithmic and computational model for Lakatosian philosophy of science [1, 12]. The underlying idea for the possibility of employing such methods in Lakatosian philosophy of mathematics is Lakatos's understanding that mathematics is a quasi-empirical activity. For Lakatos, thought-experiments reflect the empirical side of the mathematical practice. Our formalization reflects this point, too. We start from a single formula f, then extend it to a set of formula F, and finally experiment with the different formulas in F to see how they interact with the geometrical objects in question, and finally modify our set of possible worlds if necessary.

However, we are very well aware of the fact that Lakatos's vague and heavily practice based notion of heuristics cannot be fully formalized. There are several reasons for this pessimistic observation of us. One, and perhaps the most significant one is the mixed behavior of positive and negative heuristics in Lakatosian approach. There are examples and counterexamples in PR which behave both positively and negatively from a heuristics point of view [2]. This ambiguity makes it almost impossible to give a meta-mathematical account of Lakatosian heuristics, and derives us to adopt algorithm and computation based position.

We pointed out that there is a close resemblance between some game theoretical concepts and Lakatosian heuristics. Nevertheless, this connection is not well studied yet even though the basics of the game theoretical semantics of the logic we have used had been given already [1]. Thus, in our opinion, it is worthwhile to identify the Lakatosian strategies in a formal context more clearly, and we believe, under some certain restrictions and assumptions perhaps, this is a very promising research direction that can shed some more light on the behavior of heuristic methods.

What we have provided here is a new, descriptive and more rigorous way to understand Lakatosian heuristics. First of all, our contribution shows that Lakatosian heuristics is not random, and thus follows a pattern — a pattern which is perhaps almost impossible to predict beforehand. Second, we presented a formal contribution to the actual practice of mathematics. Even though Lakatos's *rationally reconstructed* presentation does not perfectly reflect the actual history of Euler's conjecture, his approach constitutes a very significant contribution to the discussions on the methodology of mathematics, and we aimed at understanding his approach from a rather formal point of view.

ACKNOWLEDGEMENT

Rohit Parikh provided an important feedback on the crucial ideas presented in this paper. Vadim Batitsky read the paper very carefully with many important remarks. Some of the central ideas of this paper have grown out of the discussions with Eric Pacuit and Samet Bağçe. I am grateful to them all. An earlier and shorter version of this paper was presented in Panhellenic Logic Symposium in 2009 in Patras, Greece. We also acknowledge the valuable feedback of the anonymous referee.

Department of Computer Science
The Graduate Center
The City University of New York
New York, USA
E-mail: cbaskent@gc.cuny.edu
Website: http://www.canbaskent.net

REFERENCES

[1] BAŞKENT, C.: *Topics in Subset Space Logic*, Masters Thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam (2007).

[2] BAĞÇE, S., BAŞKENT, C.: An Examination of Counterexamples in Proofs and Refutations, *Philosophia Scientiæ*, vol. 13, No. 2, 3–20 (2009).

[3] CORFIELD, D.: Assaying Lakatos's History and Philosophy of Science. *Studies in History and Philosophy of Science* 28, 99–121 (1997).

[4] KISS, O.: Mathematical Heuristics – Lakatos and Pólya. In Kampis, G., Kvasz, L., Stölzner, M. (eds.) *Appraising Lakatos: Mathematics, Methodology and the Man*, pp. 243–254. Kluwer, Dordrecht (2002).

[5] KISS, O.: Heuristics, Methodology or Logic of Discovery? Lakatos on Patterns of Thinking. *Perspectives on Science* 14, 302–317 (2006).

[6] KOETSIER, T.: *Lakatos' Philosophy of Mathematics: A Historical Approach.* North-Holland, Amsterdam (1991).

[7] KVASZ, L.: Lakatos' Methodology Between Logic and Dialectic. In Kampis, G., Kvasz, L., Stölzner, M. (eds.) *Appraising Lakatos: Mathematics, Methodology and the Man*, pp. 211–242. Kluwer, Dordrecht (2002).

[8] LAKATOS, I.: *Proofs and Refutations.* Cambridge University Press, Cambridge (1976).

[9] LAKATOS, I.: *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers* (1978).

[10] MOSS, L., PARIKH, R.: Topological Reasoning and The Logic of Knowledge. In Moses, Y. (ed.) *Proceedings of the 4th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK)*, pp. 95–105. Morgan Kaufmann, San Francisco (1992).

[11] PARIKH, R.: Social Software, *Synthese*, vol. 132, pp. 187–211 (2002).

[12] PEASE, A., COLTON, S., SMAILL, A., LEE, J.: *A Model of Lakatos's Philosophy of Mathematics*, Computing and Philosophy (ECAP), (2004).

[13] PRIEST, G.: *In Contradiction*, Oxford University Press (2006).

[14] YUXIN, Z.: From the Logic of Mathematical Discovery to the Methodology of Scientific Research Programmes. *British Journal of Philosophy of Science* 41, 377–399 (1990).

Appendix A. *A Quote from* Proofs and Refutations

Criticism of the Conjecture by Global Counterexample

ALPHA I have a counterexample which will falsify your first lemma. This will also be a counterexample to the main i.e. this will be a global counterexample as well. (...) Imagine a solid bounded by a pair of nested cubes — a pair of cubes, one of which is inside, but does not touch the other. (...) [F]or each cube $V - E + F = 2$, so that for the hollow cube $V - E + F = 4$.

Rejection of the counterexample. The method of monster-barring

DELTA But why accept the counterexample? We proved our conjecture — now it is a theorem. I admit that it clashes with this so-called 'counterexample'. One of them has to give way. But why should the theorem give way, when it has been proved? It is the 'criticism' that should retreat. It is fake criticism. This pair of nested cubes is not a polyhedra all. It is a *monster*, a pathological case, not a counterexample.
GAMMA Why not? A polyhedron is a solid whose surface consists of polygonal faces. And my counterexample is a solid bounded by polygonal faces.
(...)
DELTA Your definition is incorrect. A polyhedron must be a surface: it has faces, edges, vertices, it can be deformed, stretched out on a blackboard, and has nothing to do with the concept of 'solid'. A polyhedron is a surface consisting of a system of polygons.
DELTA So really you showed us two polyhedra — two surfaces, one completely inside the other. A woman with a child in her womb is not a counterexample to the thesis that human beings have one head.
ALPHA So! My counterexample has bred a new concept of polyhedron. Or do you dare to assert that by polyhedron you always meant a surface?
(...)
ALPHA (...) Take two tetrahedra which have an edge in common. Or, take two tetrahedra which have a vertex in common. Both these twins are connected, both constitute one single surface. And, you may check that for both $V - E + F = 3$.
(...)
GAMMA (...) [A] star-polyhedron — I shall call it urchin. This consists of 12 star-pentagons. It has 12 vertices, 30 edges, and 12 pentagonal faces (...). Thus the Descartes-Euler thesis is not true at all, since for this polyhedron $V - E + F = -6$.
(...)
GAMMA Do you not see? This is a polyhedron, whose faces are the twelve star-pentagons. (...)

DELTA But then you do not even know what a polygon is! A star-pentagon is certainly not a polygon!
(...)
ALPHA Consider a picture-frame like this. This is a polyhedron according to any of the definitions hitherto proposed. Nonetheless you will find,faces, that $V - E + F = 0$.
(...)
GAMMA: I have a new one (...) a simple cylinder. It has 3 faces (the top, the bottom and the jacket), 2 edges (two circles) and no vertices. (...)
DELTA Alpha stretched concepts, but you tear them! Your 'edges' are not edges! An edge has two vertices!

Improving the conjecture by exception-barring methods. Piecemeal exclusions. Strategic withdrawal or playing for safety

BETA I find some aspects of Delta's arguments silly, but I have come to believe that there is a reasonable kernel to them. It now seems to me that no conjecture is generally valid, but only valid in a certain restricted domain that excludes the exceptions. I am against dubbing these exceptions 'monsters' or 'pathological cases'. That would amount to the methodological decision not to consider these as interesting examples in their own right, worthy of a separate investigation. But I am also against the term 'counterexample'; it rightly admits them as examples on a par with the supporting examples, but somehow paints them in war-colours, so that, like Gamma, one panics when facing them, and is tempted to abandon beautiful and altogether. No: they are just exceptions.
(...)
BETA There are certainly three types of propositions: true ones, hopelessly false ones and hopefully false ones. This last type can be improved into true propositions by adding a restrictive clause which states the exceptions. I never 'attribute to formulas an undetermined domain of validity. In reality most of the formulas are true only if certain conditions are fulfilled. By determining these conditions and, of course, pinning down precisely the meaning of the terms I use, I make all uncertainty disappear.' So, as you see, I do not advocate any sort of peaceful coexistence between unimproved formulas and exceptions. I improve my formulas and turn them into perfect ones, like those in Sigma's first class. This means that I accept the method of monsterbarring in so far as it serves for finding the domain of validity of the original conjecture; I reject it in so far as it functions as 'a linguistic trick for rescuing 'nice' theorems by restrictive concepts. These two functions of Delta's method should be kept separate. I should like to baptise my method, which is characterised by the first of these functions only, 'the exception-barring method'. I shall use it to determine precisely the domain in which

the Euler conjecture holds.
(...)
BETA For all polyhedra that have no cavities (like the pair of nested cubes), tunnels (like the picture-frame), or multiple structures (twin-tetrahedra) $V - E + F = 2$.
BETA So instead of barring exceptions piecemeal, I shall draw the borderline modestly, but safely: All convex polyhedra are Eulerian. And I hope you will that this has nothing conjectural about it: that it is a theorem.
(...)

Improving the conjecture by the method of lemma-incorporation. Proof-generated theorem versus naive conjecture

ALPHA To help your imagination, I will tell you that those and only those polyhedra which you can inflate into a sphere have the property that, after a face is removed, you can stretch the remaining part onto a plane.
It is obvious that such a 'spherical' polyhedron is stretchable onto a plane after a face has been cut out; and vice versa it is equally obvious that, if a polyhedron minus a face is stretchable onto a plane, then you can bend it into a round vase which you can then cover with the missing face, thus getting a spherical polyhedron. But our pictuframe can never be inflated into a sphere; but only into a torus.
TEACHER Good. Now, unlike Delta, I accept this picture-frame as a criticism of the conjecture. I therefore discard the conjecture in its original form as false, but I immediately put forward a modified, restricted version, namely this: the Descartes-Euler conjecture holds good for 'simple' polyhedra, i.e. for those which, after having had a face removed, can be stretched onto a plane. Thus we have rescued some of the original hypothesis. We have: The Euler characteristic of a simple polyhedron is 2. This thesis will not be falsified by the nested cube, by the twin-tetrahedra, or by star-polyhedra — for none of these is 'simple.'
So while the exception-barring method restricted both the domain the main conjecture and of the guilty lemma to a common domain of safety, thereby accepting the counterexample as criticism both of the main conjecture and of the proof, my method of lemma-incorporation upholds the proof but reduces the domain of the main conjecture to the very domain of the guilty lemma. Or, while a counterexample which is both global and local made the exception-barrer revise both the lemmas and the original conjecture, it makes me revise the original conjecture, but not the lemmas.
(...)

The method of proof and refutations

LAMBDA Let me state its main aspects in three heuristic rules:
Rule I. If you have a conjecture, set out to prove it and to refute it. Inspect the proof carefully to prepare a list of non-trivial lemmas (proof-analysis); find counterexamples both to the conjecture (global counterexamples) and to the suspect lemmas (local counterexamples).
Rule 2. If you have a global counterexample discard your conjecture, add to your proof-analysis a suitable lemma that will be refuted by it, and replace the discarded conjecture by an improved one that incorporates that lemma as a condition. Do not allow a refutation to be dismissed as a monster. Make all 'hidden lemmas' explicit.
Rule 3. If you have a local counterexample, check to see whether it is not also a global counterexample. If it is, you can easily apply Rule 2.