



HAL
open science

Liberté et négation. Ceci n'est pas un festschrift pour Imre Toth.

Siegmund Probst, Agnès Erdélyi, Antonio Moretto, Karine Chemla

► **To cite this version:**

Siegmund Probst, Agnès Erdélyi, Antonio Moretto, Karine Chemla. Liberté et négation. Ceci n'est pas un festschrift pour Imre Toth.. editors Siegmund Probst and Karine Chemla with the collaboration of Agnès Erdélyi and Antonio Moretto, 1996. halshs-00004274v2

HAL Id: halshs-00004274

<https://shs.hal.science/halshs-00004274v2>

Submitted on 30 Aug 2005 (v2), last revised 20 Dec 2005 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Withdrawing hypotheses using negation as failure

Lorenzo Magnani

Computational Philosophy Laboratory

Department of Philosophy

University of Pavia

I-27100 Pavia – Italy

lorenzo@philos.unipv.it

Abstract

Computer and AI scientists have suggested an interesting technique for negating hypotheses and accessing to the new ones: negation as failure. The general objective is to consider how the use of *negation as failure* may be relevant to *hypothesis withdrawing*. There has been little research into the weak kinds of negating hypotheses, despite abundant reports that hypothesis withdrawing is crucial in everyday life but also in certain kinds of diagnostic or epistemological settings, such as medical reasoning and scientific discovery. In the cases of conceptual change I describe inferences are made using such kind of negation, as a fundamental tool for advancing knowledge: new conclusions are issued on data responsible of the failure of the previous ones. To delineate some of the features of what I call *abducers*, I plan to explore whether this kind of negation can be employed to model hypothesis withdrawing in Freudian analytic reasoning and in Poincaré's conventionalism of the first principles of physics.

1 Change in theoretical systems

Contradiction and its reconciliation play an important role in philosophy, in scientific theories and in all kinds of problem-solving. It is the driving force underlying change (thesis, antithesis and synthesis) in the Hegelian dialectic and the main tool for advancing knowledge (conjectures and refutations [15] and proofs and counter-examples [7]) in the Popperian philosophy of science and mathematics.

Following Quine's point of view against the distinction between necessary and contingent truths [16], when a contradiction arises, consistency can be

restored by rejecting or modifying any assumption which contributes to the derivation of contradiction: no hypothesis is immune from possible alteration. Of course there are epistemological and pragmatic limitations: some hypotheses contribute to the derivation of useful consequences more often than others, and some participate more often in the derivation of contradictions. For example it could be useful to abandon, among the hypotheses which lead to contradiction, the one which contributes least to the derivation of useful consequences; if contradictions continue to be and assessed utility of the hypotheses changes, it may be necessary to backtrack, reinstate a previously abandoned hypothesis and abandon an alternative instead.

Hence, the derivation of inconsistency contributes to the search of alternative hypotheses, may be new: for each assumption which contributes to the derivation of a contradiction there exists at least one alternative new system obtained by abandoning or modifying the assumption.

The classical example in which a theoretical system is in front of a contradiction is the case in which the report of an empirical observation or experiment contradicts a scientific theory. Whether it is more beneficial to reject the report or a statement of a theory depends on the whole effect on the theoretical system. It is also possible that many alternatives might lead to incomparable, equally viable, but mutually incompatible, systems¹.

As argued by Lakatos, in a mature theory with a history of useful consequences, it is generally better to reject an anomalous conflicting report than it is to abandon the theory as a whole. The cases in which we have to abandon a whole theory are very rare: a theory may be considered as a complex information system in which there is a collection of cooperating individual statements some of which are useful and more firmly held than others; propositions that belong to the central core of a theory are more firmly held than those which are located closer to the border, where instead rival hypotheses may coexist as mutually incompatible alternative. Accumulating reports of empirical observations can help in deciding in favor of one alternative over another.

We have to remember that even without restoring consistency, an inconsistent system can still produce useful information. Of course from the point

¹Thagard proposes a very interesting computational account of scientific controversies in terms of the so-called *explanatory coherence* [19], that improves the classical Lakatos'one [8]. Also related to the problem of the so-called "belief change" is Levi's theory of suppositional reasoning [9].

of view of classical logic we are compelled to derive from inconsistent premises any conclusion, but in practice efficient proof procedures infer only “relevant” conclusions with varying degrees of accessibility, as stated by the criteria of non-classical *relevant entailment* [1].

We can conclude asserting that contradiction, far from damaging a system, helps to indicate regions in which it can be changed (and improved). Contradiction has a preference for hard hypotheses which are more easily falsified than timid ones; what’s more, hard hypotheses may more easily weakened than timid ones, which are difficult to strengthen later on. It is always better to produce mistakes and to correct them than to make any progress.

Let us consider now a kind of “timid” hypothesis that is hard to negate and the ways for making it easily. In these cases the subject can *rationally* decide of withdrawing his hypotheses also in contexts where it is *impossible* to find “*explicit*” contradictions; more than that, thank to the new information reached just in finding this kind of negation, the subject becomes *free* of guessing new hypotheses. I will explore whether negation as failure can be employed to model hypothesis withdrawing in Freudian analytic reasoning and in Poincaré’s conventionalism of the first principles of physics (solving respectively the questioned problem of the probative value of clinical findings and showing how conventions can be motivationally abandoned)

2 Negation as failure in query evaluation

There is a kind of negation, found by computer science researchers, that I consider very important from the epistemological point of view: *negation as failure*. It is active as a “rational” process of withdrawing guessed hypotheses in everyday life but also in certain subtle kinds of diagnostic and epistemological settings. Contrasted with classical negation, with the double negation of intuitionistic logic, and with the philosophical concept of *Aufhebung*, negation as failure shows how a subject can decide of withdrawing his hypotheses, maintaining the *rationality* of his reasoning, in contexts where it is impossible to find contradictions; as stated above, thank to the new information reached just in finding this kind of negation, the subject becomes *free* of forming new hypotheses.

Let us consider a special query evaluation process for a logic data base (comprising a set of Horn clauses) that involves the so-called *negation as failure*

inference rule [2]. We can build a Horn clause theorem prover augmented with this special inference rule. By this inference rule we are able of inferring $\neg P$ when every possible proof of P fails.

We know that a relational data base only contains information about *true* instances of relations. Even so, many queries involve negation and we can answer them by showing that certain instances are *false*. For example let's consider this simple case: to answer a request for the name of a student not taking a particular course, C , we need to find a student, S , such that the instance (atomic formula) $Takes(S, C)$ is false. For a logic data base, where an atomic formula which is not explicitly given may still be implied by a general rule, the assumption is that an atomic formula is false if we *fail* to prove that it is true. To prove that an atomic formula P is *false* we do an exhaustive search for a proof of P . If *every* possible proof of P fails, we can infer $\neg P$. The well-known PROLOG programming language [17] uses this kind of manipulating negation.

We have to deal with a proof like the following:

- from proving $\not\vdash P$ infer $\vdash \neg P$

where the “proof that P is not provable” ([2], p. 120) is the *exhaustive* but *unsuccessful search* for a proof of P . Here the logical symbol \neg acquires the new meaning of “fail to prove”. What is the semantic significance of this kind of negation? Can we interpret a failure proof of $\neg P$ as a *valid* first order inference that P is false? [to be completed]

In the following sections I will show how it is possible to explain the epistemological status of Freud's method of clinical investigation in terms of negations as failure; we will also consider Poincaré's famous conventionalism of the first principles of physics.

3 Withdrawing “constructions”

We are not dealing here with the highly controversial problem of the epistemological status of the psychoanalytic clinical theories²: it is well-known that clinical data have no probative value for the confirmation or falsification

²This problem is fully analysed in [5].

of the hypotheses of psychoanalytic clinical theories of personality: because completely involved in the specificity of clinical setting, they are devoid of the independence that characterizes observations endowed with scientific value. What's more, because of the lacking of probative value, any therapeutic gains from analysis may be considered as not wrought by true insightful self-discovery but rather are placebo effects induced by the analyst's suggestive influence. If the probative value of the analysand's responses is negated, then Freudian therapy might reasonably be held to function as an emotional corrective (carried by a positive "transference" effect) not because it enables the analysand to acquire self-knowledge, but instead because he or she succumbs to proselytizing *suggestion*, which operates the more insidiously under the pretense that analysis is nondirective. The suggestion is indeed responsible of the so-called epistemical contamination of the patient's responses.

Freud asks the patient to believe in the analyst's theoretical retrodictions of significant event in his early life and these theoretical retrodictions are communicated to him as *constructions*. The aim is to provoke the previously cited true insightful self-discovery that guarantees the cure ([3] 1920, vol. 18, p. 18). A singular construction is built as a "sequence" of the *interpretations* issued from clinical data found in the clinical setting, epistemologically characterized by "transference" and "countertransference". A construction can be considered as a kind of "history" or "narrative" of the analysand's significant early life happenings, never complete, but more and more extensible adding new interpretations.

Freudian clinical reasoning refers to a kind of *abductive reasoning* I called *selective* [12]: its uncertainty is due to *nonmonotonicity*, the analyst may always withdrawn his or her interpretations (constructions) in front of new evidences. Every "abduced" construction, suitably connected with some other clinical psychoanalytical hypotheses, generates expectations with regard to the analysand's subsequent responses and remarks.

Of course the analyst aims to build *the most complete* construction. The problem here is the analyst cannot propose to the analysand the construction he wants, without any external testing. As stated above, the objection most often raised against psychoanalysis is that "therapeutic success in *nonprobative* because it is achieved *not* by imparting veridical insight but rather by the persuasive suggestion of fanciful psudoinights that merely ring verisimilar to the docile patient" ([5], p. 138). In his very later paper *Constructions in*

analysis ([3] 1937, 23, pp. 257-269) Freud reports that “a certain well-known man of science” had been “at once derogatory and unjust” because

He said that in giving interpretations to a patient we treat him upon the famous principle of “Heads I win, tails you lose” [footnote omitted]. That is to say, if the patient agrees with us, then the interpretation is right, but if he contradicts us, that is only a sign of his resistance, which again shows that we are right. In this way we are always in the right against the poor helpless wretch whom we are analysing, no matter how he may respond to what we put forward ([3] 1937, 23, p. 257).

Freud is looking for a criterion for justifying, in the clinical setting, the abandonment of constructions that have been shown as inadequate³. This is the fundamental epistemological problem of the method of clinical investigation: Freud is clear in saying that the therapeutic success will occur only if incorrect analytic constructions, spuriously confirmed by “contaminated” responses from the patient, have been discarded in favor of new correct constructions (that are constitutively *provisional*) derived from clinical data not distorted by the patient’s compliance with the analyst’s communicated expectations.

Freud therefore proceeds “to give a detailed account of how we are accustomed to arrive at an assessment of the ‘Yes’ and ‘No’ (considered as “direct evidences”) of our patients during analytic treatment - of their expression of agreement or of denial” (p. 257).

Analytic constructions cannot be falsified by dissent from the patient because “it is in fact true that a ‘No’ from one of our patients is not as a rule enough to make us abandon an interpretation as incorrect” (p. 257). It might seem to Freud that patient dissent from an interpretation can be always discounted as inspired by neurotic resistance. It is only “in some rare cases” that dissent “turns out to be the expression of legitimate dissent” (p. 263). A “patient’s ‘No’ is no evidence of the incorrectness of a construction, though it is perfectly compatible with it”. Rather, patient’s ‘No’ might be more adequately related to the “incompleteness” of the prospected constructions: “the only safe interpretation of his ‘No’ is that it points to incompleteness” (p. 263).

Even if a patient’s verbal assent may result from genuine recognition that the analyst’s construction is true, it may be spurious by deriving from neurotic

³It is interesting to note that in the cited article Freud emphasizes the provisional role of constructions referring to them also as “hypotheses” or “conjectures”.

resistance, as already seen in his or her dissent. Assent is “hypocritical” when serves “to prolong the concealment of a truth that has not been discovered”. The assent is genuine and not hypocritical when patient’s verbal assent will be followed and accompanied by new memories as flows: “The ‘Yes’ has no value unless it is followed by indirect confirmations, unless the patient, immediately after his ‘Yes’, produces new memories which complete and extend the construction” (p. 262)

Since “Yes” and “No” do not have any importance to test a construction it is necessary to see other facts, such as “the material that crops out” after having proposed to the patient a construction:

[...] what happens [...] is that the patient remains as impassive, and does not react neither answering ‘Yes’ nor answering ‘No’ to the prospected construction. May be this condition signifies a postponement of the reaction, nevertheless, if the condition does not change, it is legitimate to conclude we made a mistake. When convenient we will admit it with the patient without any detriment of our authority (p. 261).

Let us analyse now this situation from the epistemological point of view: the analyst has to withdraw the construction when has failed to prove the construction (hypothesis). Remember that for a logic data base the assumption is that an atomic formula is false if we *fail* to prove that it is true. More clearly: as stated above, every construction, suitably connected with some other clinical psychoanalytical hypotheses, generates expectations with regard to the analysand’s subsequent responses and remarks. We consider as a proof of a construction the fact that we can more and more extend and complete it adding the new (*expected*) material that “crops out” from the patient. If the patient does not show this new “material” able to extend the prospected construction, this *failure* leads to withdraw the construction itself. So the “proof that a construction is not provable” is the *unsuccessful search* for a proof of the construction itself. After a finite time Here the logical symbol \neg acquires the new meaning of “fail to prove” in the empirical sense. [to be completed]

4 Withdrawing “conventions”

When Poincaré illustrates the problem of abandoning conventional principles, he is dealing with a kind of timid negation:

It is the mathematical physics of our fathers which has familiarized us little by little with these various principles [conventional]; which has habituated us to recognize them under the different vestments in which they disguise themselves ([14], p. 95).

and then

Have you not written, you might say if you wished to seek a quarrel with me - have you not written that the principles, though of experimental origin, are now unassailable by experiment because they have become conventions? And now you have just told us that the most recent conquests of experiment put these principles in danger.

Well, formerly I was right and to-day I am not wrong. Formerly I was right, and what is now happening is a new proof of it. Take, for example, the calorimetric experiment of Curie on radium. Is it possible to reconcile it with the principle of the conservation of energy? This has been attempted in many ways; but there is among them one I should like you to notice; this is not the explanation which tends to-day to prevail, but it is one of those which have been proposed. It has been conjectured that radium was only an intermediary, that it only stored radiations of unknown nature which flashed through (cambio pagina) space in every direction, traversing all bodies, save radium, without being altered by this passage and without exercising any action upon them. Radium alone took from them a little of their energy and afterward gave it out to us in various forms.

What an advantageous explanation, and how convenient! First, it is unverifiable and thus irrefutable. Then again it will serve to account for any derogation whatever to Mayer's principle; it answers in advance not only the objection of Curie, but all the objections that future experimenters might accumulate. This new and unknown energy would serve for everything.

This is just what I said, and therewith we are shown that our principle is unassailable by experiment.

But then, what have we gained by this stroke? The principle is intact, but thenceforth of what use is it? It enabled us to foresee that in such and such a circumstance we could count on such a total quantity of energy; it limited us; but now that this indefinite provision of new energy is placed at our disposal, we

are no longer limited by anything; and, as I have written in “Science and Hypothesis”, if a principle ceases to be fecund, experiment without contradicting it directly will nevertheless have condemned it ([14], pp.109-110).

[to be completed]

References

- [1] Anderson, A. & Belnap, N., *Entailment*, Princeton University Press, Princeton, 1975.
- [2] Clark, K.L., *Negation as failure*, in H. Gallaire & J. Minker (eds.), *Logic and Data Bases*, Plenum, New York, 1978, pp. 119-140. (Reprinted in [4], pp. 311-325).
- [3] Freud, S., *Standard Edition of the Complete Psychological Works of Sigmund Freud*, translated by J. Strachey et al., Hogarth Press, London, 1953-1974.
- [4] Ginsberg, M.L. (ed.), *Readings in Nonmonotonic Reasoning*, Morgan Kaufmann, Los Altos, CA, 1987.
- [5] Grünbaum, A., *The Foundations of Psychoanalysis. A Philosophical Critique*, University of California Press, Berkeley and Los Angeles, CA, 1984.
- [6] Kowalski, R., *Logic for Problem Solving*, North Holland, Amsterdam, 1979.
- [7] Lakatos, I., *Proofs and Refutations. The Logic of Mathematical Discovery*, Cambridge University Press, Cambridge, 1976.
- [8] Lakatos, I., History of science and its rational reconstructions, in R. Buck & R.S. Cohen (eds.), *PSA 1970: In memory of Rudolf Carnap*, Reidel, Dordrecht, 1971.
- [9] Levi, I., *For the Sake of the Argument. Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning*, Cambridge University Press, Cambridge, 1996.
- [10] Magnani, L., *Epistemologia applicata*, Marcos y Marcos, Milano, 1991.

- [11] Magnani, L. (ed.), *Conoscenza e matematica*, Marcos y Marcos, Milano, 1991.
- [12] Magnani, L., *Abductive reasoning: philosophical and educational perspectives in medicine*, in D.A. Evans and V.L. Patel (eds.), *Advanced models of cognition for medical training and practice* Springer, Berlin, 1992, pp. 21-41.
- [13] Poincaré, H., *La science et l'hypothèse*, Flammarion, Paris, 1902, 1968, 1970 (English translation by G.B.Halsted, *The Value of Science*, Dover Publications, New York, 1958).
- [14] Poincaré, H., *La valeur de la science*, Flammarion, Paris, 1905, 1909, 1970 (English translation by G.B.Halsted, *The Value of Science*, Dover Publications, New York, 1958).
- [15] Popper, K., *Conjectures and Refutations. The Growth of Scientific Knowledge*, Routledge and Kegan Paul, London, 1963.
- [16] Quine, W.V.O, Two dogmas of empiricism, *Philosophical Review*, 40, 1951, 113-127. Also in Quine, W.V.O., *From a Logical Point of View*, Hutchinson, London, 1953, 1961², pp. 20-46.
- [17] Roussel, P., *PROLOG: Manual d'Utilisation*, Rapport interne, G.I.A., UER de Luminy, Université d'Aix-Marseille, 1975.
- [18] Sombé, L., Reasoning under incomplete information in artificial intelligence. A comparison of formalisms using a single example, *International Journal of Intelligent Systems*, 5(4), 1990, Special Issue.
- [19] Thagard, P., *Conceptual Revolutions*, Princeton University Press, Princeton, 1992.
- [20] Toth, I., Essere e non essere: il teorema induttivo di Saccheri e la sua rilevanza ontologica, translated by A. Marini, in [11], pp. 87-156.