Computational 'Consilience' as a Basis for Theory Formation

José Hernández-Orallo[1]
Universitat Politècnica de València
Departament de Sistemes Informàtics i Computació,
Camí de Vera 14, Aptat. 22.012 E-46071, València, Spain
E-mail: `jorallo@dsic.upv.es`.

Extended Abstract:

We present computational definitions for the notion of 'consilience' of a theory. The term 'consilience' was coined by Whewell [Whewell 1847] to comprise the relevant basics in scientific theories: prediction, explanation and unification of fields. Since all of these criteria are desirable, consilience was *informally* introduced as a foundational issue for theory construction and modelling. However, a unified, formal and computational conception has not been presented to date, integrating in a consistent way prediction, explanation and unification of fields, allowing the growth and revision of knowledge.

Nonetheless, the essence of our definition is based on a very well known psychological ground for ontology and epistemology: the notion of reinforcement. Whatever the approach to knowledge construction, the construction or revision of knowledge must come from a gain or loss, respectively, of reinforcement (also known as apportionment of credit [Holland et al. 1986]). We present a way to compute the reinforcement map for a given theory, depending on past observations.

The usual or pure reinforcement $\rho\rho_T(r)$ of a rule r from a theory T wrt. to some given observation $C = \{c_1, c_2, \ldots, c_n\}$ is computed as the number of proofs for c_i where r is used. If there are more than one proof for a given c_i, *all* of them are reckoned, but, in the same proof, a rule is computed only once. The (normalised) reinforcement $\rho_T(r)$ is defined as

$$\rho_T(r) = 1 - 2^{-\rho\rho(r)}.$$

From these definitions some properties are proven. For instance, in general, the most reinforced theory is not the shortest one but, *in the limit*, simplicity is a good criterion to obtain consilience. Somehow surprisingly, this is not the case in

1 Also at the Department of Logic and Philosophy of Science of the University of Valencia. On-line papers: `http://www.dsic.upv.es/~jorallo/escrits/ escritsa.htm`.

finite situations, and even some kind of redundancy (*investment*) does not necessarily imply a loss of reinforcement ratio.

However, this measure of reinforcement of the *theory* suffers the appearance of *fantastic* concepts. The rationale relies on the fact that an invented rule, used in every other rule of the theory, could *unjustifiably* increase the reinforcement ratio of a theory. Although a simplicity criterion can be used to avoid these fantastic concepts, it would make our proposal lose some of its interesting properties. Fortunately, our peculiar way out comes by measuring the validation *wrt. the data*:

The *course* $\chi_T(f)$ of a given fact f wrt. to a theory T is computed as the product of all the reinforcements $\rho(r)$ of all the rules r used in the proof of f. If a rule is used more than once, it is computed once. If f has more than one proof, the greatest course is selected.

With this definition, it is proven that no fantastic rule can be added in the previous way, but the good properties of the original definition are still preserved.

From here, one has the dilemma of selecting the theory with the greatest *mean* of the courses of all the data presented so far, or maybe one wants a compensated theory, where a *geometric mean* can be used instead. Nevertheless, in explanatory induction, the theory must explain all the data, i.e., it cannot have any anomaly. Consequently, one would discard theories where a fact has a course value less than the mean divided by a *strictness* constant. It is shown that this *strictness* constant corresponds to the degree of consilience we were looking for.

The definitions are applied and illustrated with some examples of knowledge construction and revision, using first-order logic as representation language. Some theories are generated from the evidence using both descriptional induction (with no required consilience at all) and explanatory induction (with high consilience). The examples show that, only in the latter case, abduction can be naturally incorporated as a special case of explanatory induction. In abduction, new *factual* hypotheses are added to the theory in order to increase or maintain consilience. Obviously, this works as long as no anomaly is accepted to the theory; in misspelled words, the new evidence (or novelty) must be *consiliated* by the theory, but never quoted extensionally.

In this way, our differentiation between Enumerative (or descriptional) Induction and Best Explanation [Ernis 1968] [Harman 1965] [Hempel 1965] (see [Sharger & Langley 1990] for more modern contrasted positions in this debate) is not based on the predictive value of a theory but on the *usability* (in terms of a coherent explanation of reality) and the applicability of abduction.

Moreover, in explanatory induction, deduction *can* and *must* play a very important role. The relation between consilience and the modern view of "explanatory coherence" [Thagard 1998] as *deductive* constraint satisfaction is investigated. It seems that the combination of both represents the traditional notion of 'coherence' of scientific theories [Thagard 1978], [Thagard 1989]. Summing up, consilience ensures that the theory is justified by the evidence and coherence ensures that the theory is the most *compatible* with the background knowledge constructed from other evidences. The clearest case happens when a compatibility or satisfiability check for coherence shows that some explanation H is inconsistent with a previous knowledge T. In this case, the reinforcement of H wrt. to the new evidence E, i.e. $\chi_H(E)$, must be compared with the reinforcement of the past evidence which is in conflict with H. As a result, the usual alternatives are: the new hypothesis is discarded or the background theory revised (and perhaps in this process a *consilient* new theory can be found, i.e. an explanation)[2].

Finally, deduction must not be longer seen as a static and non-creative process that does not bring any *information*. In addition, induction and abduction should not be seen as inverse processes of deduction, *in terms of information gain*. Indeed, any computational induction and abduction must be done in a computational system, so it is deductive somehow[3]. Descriptional induction denies that deduction can increase information (although theorem proving, for instance, is a very informative field). In this point, further work is under development to reconcile deduction, induction and abduction [Hernandez-Orallo 1998].

A last question deals with the choice of a reasoning process that can make consilient theories. We formally state that analogy favours consilience. Moreover, analogy is confirmed as the fundamental mechanism for obtaining consilient theories. The reason is simple: analogy extracts a common superstructure between two situations, and this 'shared' superstructure is reinforced by both situations but, as we have seen, only grounded concepts and not 'forced' fantastic ones are allowed as valid analogies.

In conclusion, consilience is the key for many kinds of explanatory induction (including analogical reasoning and abduction) and it can be seen as a basis for the growth of knowledge and theory revision. Many important traits are quite remarkable over descriptional induction: not only induction and abduction are informative processes; deduction can increase information (like [Hintikka 1970] advocated), without making the whole theory less likely. Therefore the classical view of probability and information inversely related by $P(h) = 2^{-I(h)}$ [Bar-Hillel & Carnap 1953] is neglected.

Keywords: Consilience, Model and Theory Formation, Explanatory Induction, Abduction, Reinforcement, Analogy, Coherence, Philosophy of Science, Deduction and Information.

References

[Barker 1957] Barker, S.F. *Induction and Hypothesis* Ithaca, 1957.

[2] Note that under consilience and coherence considerations, revision is much more frequent than extension.

[3] The idea backs to the introduction in 1949 of the deductive-nomological model of explanation by Hempel and Oppenheim [Hempel 1965], which argued that abduction is just a selection of possible phenomena derived from very general laws (*nomos*).

[Bar-Hillel & Carnap 1953] Bar-Hillel, Y.; Carnap, R. "Semantic Information" *British Journal for the Philosophy of Science* 4, 1953, 147-157.

[Boutilier & Becher 1995] Botilier, C.; Becher, V. "Abduction as belief revision" Artificial Intelligence 77, 43-94, 1995.

[Carnap 1952] Carnap, R. "The Continuum of Inductive Methods" University of Chicago, Chicago, IL, 1952.

[Dieterich & Flann 1997] Dietterich, T.G.; Flann, N.S. "Explanation-Based Learning and Reinforcement Learning: A Unified View" Machine Learning, 28, 169-210, 1997.

[Ernis 1968] Ernis, R. "Enumerative Induction and Best Explanation" The Journal of Philosophy, LXV (18), 523-529, 1968.

[Flach 1996] Flach, P. "Abduction and Induction: Syllogistic and Inferential Perspectives" in M. Denecker, L. De Raedt, P. Flach and T. Kakas (eds) Working Notes of the ECAI'96 Workshop on *Abductive and Inductive Reasoning*, pp. 7-9, 1996.

[Flach & Kakas 1998] Flach, P.; Kakas, A. (eds) "Abduction and Induction. Essays on their relation and integration", in press, Kluwer 1998.

[Kling 1971] Kling, R.E. "A Paradigm for reasoning by analogy" Artificial Intelligence, 2:147:178, 1971.

[Harman 1965] Harman, G. "The inference to the best explanation" *Philosophical Review*, 74, 88-95, 1965.

[Hempel 1965] Hempel, C.G. "Aspects of Scientific Explanation" The Free Press, New York, N.Y. 1965.

[Hernández-Orallo 1998] Hernández-Orallo, J. "Explanatory Induction and Informative Deduction using Intensional Complexity" Thesis Dissertation, Department of Logic and Philosophy of Science, Universitat de València, 1998, forthcoming.

[Hernandez-Orallo & Garcia-Varea 1998a] "Distinguishing Abduction and Induction under Intensional Complexity" in Flach, P.; Kakas, A. Proceedings of the ECAI'98 Workshop on Abduction and Induction in AI, August 1998, to appear.

[Hernandez-Orallo & Ramirez-Quintana 1998a] "Inductive Inference of Functional Logic Programs" J. Lloyd (ed) CompulogNet Area Meeting on Computational Logic and Machine Learning at the 1998 Joint International Conference and Symposium on Logic Programming (JICSLP'98), pp. 49-55, 1998.

[Hernandez-Orallo & Ramirez-Quintana 1998b] "Inverse Narrowing for IFLP" in Freire-Nistal, J.L.; Falaschi, M.; Vilares-Ferro, M. (eds) Proceedings of the 1998 Joint Conference of Declarative Programming, APPIA-GULP-PRODE'98, pp. 379-392, 1998.

[Hintikka 1970] Hintikka, J. "Surface Information and Depth Information" in Hintikka, J.; Suppes, P. (eds) D. Reidel Publishing Company, pp. 263-297, 1970.

[Holland et al. 1986] Holland, J.H.; Holyoak, K.J.; Nisbett, R.E.; Thagard, P.R. "Induction. Processes of Inference, Learning and Discovery" The MIT Press 1986.

[Kaelbling et al. 1996] Kaelbling, L.; Littman, M.; Moore, A. “Reinforcement Learning: A survey” *Journal of Artificial Intelligence Research*, 4: 237-285, 1996.

[Karmiloff-Smith 1992] Karmiloff-Smith, A. “Beyond Modularity: A Developmental Prespective on Cognitive Science” The MIT Press 1992.

[Kuhn 1970] Kuhn, T.S. “The Structure of Scientific Revolutions”, University of Chigago 1970.

[Muggleton & De Raedt 1994] Muggleton, S. & De Raedt L. “Inductive Logic Programming — theory and methods” Journal of Logic Programming, 19-20:629-679, 1994.

[Ng & Mooney 1990] Ng.H.; Mooney, R. "On the role of coherence in abductive explanation" in *Proceedings of the Eighth National Conference on Artificial Intelligence*, pp. 337-342 Boston, MA. AAAI, 1990.

[Peirce 1867/1960] Peirce, C.S. "Collected papers of Charles Sanders Peirce" Cambridge. Harvard University Press 1960.

[Popper 1962] Popper, K.R. “Conjectures and Refutations: The Growth of Scientific Knowledge” Basic Books, New York 1962.

[Sharger & Langley 1990] Shrager, J.; Langley, P. “Computational Models of Scientific Discovery and Theory Formation” Morgan Kaufmman, 1990.

[Thagard 1978] Thagard, P. “The best explanation: Criteria for theory choice” Journal of Philosophy, 75, 76-92, 1978.

[Thagard 1989] Thagard, P. “Explanatory coherence” *The Behavioural and Brain Sciences*, 12 (3), 435-502, 1989.

[Thagard 1998] Thagard, P. “Probabilistic networks and explanatory coherence” in P. O’Rorke & J. Josephson (eds), Automated Abduction: Inference to the best explanation, Menlo Park, AAAI Press 1998.

[Whewell 1847] Whewell W. “The Philosophy of the Inductive Sciences” New York: Johnson Reprint Corp. 1847.

[Winston 1992] Winston P.H. “Artificial Intelligence” Third Edition, Addison-Wesey Publishing Company 1992.