On “Model-based” Abduction

Jaap Kamps

Institute for Logic, Language and Computation
University of Amsterdam, Email: ⟨kamps@ccsom.uva.nl⟩

Abstract

This paper reports on a concerted effort to axiomatize social science theories in first order logic. Most social science theories are stated in ordinary language (like essay-style articles in social science journals). The natural language argumentation is usually sketchy and incomplete, relying on the reader’s common-sense or on familiarity with common background assumptions in the substantive field at hand. As a result of this, it is more than likely that some of the informal claims cannot be rigorously proved in an initial formal rendition of an ordinary language theory. This can be established formally by generating one or more counterexamples to a particular conjecture—that is, if we can find models of the premises in which the conjecture is false, we have proved that the conjecture is not a theorem. As it turns out, inspecting these models that are counterexamples to a particular conjecture can be instrumental in deciding how to revise the initial formal rendition of a theory. This suggests a “model-based” form of abduction in which revisions of the formal theory are made to account for particular sets of models.

The axiomatization of a scientific theory is traditionally viewed as the final step in the justification of a theory. A first order logic rendition of a theory gives an explicit, unambiguous exposition of the theory. This allows, in turn, for assessing various logical properties of a formalized theory: such as consistency, soundness of derivations, satisfiability and falsifiability of theorems, and so on. Establishing these theoretical criteria amounts to finding a particular proof or model, a task which can be greatly facilitated by generic tools from the domain of automated reasoning [Kamps, 1998, 1999]. Standard epistemology requires that theories be consistent, that theoretical conjectures can be soundly derived from an explicit set of axioms, etc. We can assess the status of a theory by determining these properties, making them criteria for evaluating scientific theories. Traditionally, these logical properties or criteria are viewed as a rigid, final tests. However, these criteria turn out to be especially useful during the *process* of formalizing a theory. The criteria can provide useful feedback on how to revise a formal theory in case of a deficiency. For example, they suggest how to improve an initial formal rendition of a theory by identifying implicit (background) assumptions that are taken for granted in the informal version of the theory. Ultimately, this should lead to a formal rendering that resonates closely with the ordinary language exposition of the theory.

This paper reports on a concerted effort to axiomatize social science theories in first order logic. These efforts have already resulted in various first order logic versions of scientific theories originally stated in ordinary language [for example, see Péli et al., 1994, Hannan, 1998, Kamps and Pólos, 1999]. The main obstacle for logical formalizing such a discursive theory is their rational reconstruction: interpreting the text, distinguishing important claims and argumentation from other parts of the text, and reconstructing the argumentation. This reconstruction is seldom a straightforward process (although there are some useful guidelines [Fisher, 1988]). When the theoretical statements are singled-out, they can be formulated in first order logic. This initial formal rendition can be evaluated by criteria such as consistency, and soundness of arguments. A strict justification point of view would stop after evaluating the theory by these criteria. However, since theories stated in ordinary language are typically partial and incomplete, it is highly unlikely that our initial formal version of the theory is completely satisfactory. For example, the initial formal rendition may turn out to be inconsistent, or some of the theory's claims may not be derivable. Finding such undesirable properties does reveal certain deficiencies of our initial formal rendition of the theory. Is it justified to pass this verdict on to the original theory? It seems not, and as a result, we attempt to revise our initial formal version such that it is a better reconstruction of the original theory. Fortunately, analyzing the criteria can provide useful feedback for the revision of our initial axiomatization. Of course, it may be the case that an informal claim turns eventually out to be a false conjecture. More frequently there is another explanation: authors typically assume a body of common background knowledge. It is therefore unlikely that all needed premises are explicitly mentioned in the source text. One of the thorniest problems during the logical axiomatization of an ordinary language theory, is to identify the existence and necessity of these implicit background assumptions and find ways to make them explicit. Based on only the explicit assumptions of the text, it is likely that some of the theory's claims turn out to be underivable. That is, some of the theory's claims simply do not hold in our initial formal rendition of the theory.

In order to establish formally that there is no sound derivation of an informal claim or conjecture, we have to find one or more counterexamples to a particular conjecture—that is, if we can find models of the premises in which the conjecture is false, we have proved that the conjecture is not a theorem. This can be done using an automated model generator (a computer program that can enumerate small models of a set of sentences).[1] Examining these models that are counterexamples gives crucial feedback for deciding between possible revisions of the theory. Based on the inspection of the models that are counterexamples, we can distinguish between four different cases:

Nonintended model It may be the case that, although formally a model of the set of premises, it is a non-intended model in the sense that we did not intend this model to belong to the theory—in short, the model is "erratic" in a certain respect. For example, the model may conflict with our common sense, or with implicit background knowledge in the domain of the theory. In this case we have to add explicitly this common sense or background knowledge to the premises of the theory. That is, the aim is to formulate an additional premise that is false in the model at hand. Next, we can try anew if the conjecture is now derivable. This is not necessarily the case since there may be more implicit assumptions of the theory, giving rise to different counterexamples.

Overstated conjecture The model may be a intended model of the theory and in this model the conjecture is false—it is a genuine counterexample. In this case, the original conjecture is too strong to be supported by the premises, and we have to modify it. A typical example is the case in which the conjecture is overstated, and the model presents an exception that should be excluded. After weakening the conjecture such that it holds in this model, we can test again whether this weaker conjecture is now derivable. There may be other exceptions to the conjecture, which have to be dealt with sequentially. Note that strictly speaking the theory (being the deductive closure of the set of premises) does not change when we reformulate one of the conjectures.

Underrestricted domain The model may also reveal that the conjecture of the original theory only holds on a restricted domain. This can be due to a hiatus in the original argumentation, for example in case a

[1]In many cases there do exist small models that can be generated using a model generator. There are however important limitations on the use of these programs both in principle (first-order logic is undecidable) and in practice (time, memory, and CPU).

crucial assumption has been overlooked. In this case, we will need to add further assumptions that will restrict the theory's domain. The aim is to formulate additional premisses that are false in the models that are counterexamples. Next, we can try anew if the conjecture is now derivable from the extended set of premises.

False conjecture Of course, it may also be the case that the none of the possible revisions is acceptable, and that the conjecture has to be retracted altogether.

This approach to theory revision by repeatedly considering sets of counterexamples is essentially interactive. Human theorizers can use their understanding of the domain for identifying whether a counterexample is a non-intended model or a genuine counterexample representing an overlooked exception. These distinctions are crucial because they distinguish between possible revision of the premises (in case of a non-intended model), or whether we need to revise the conjecture (in case of a genuine counterexample). This decision is impossible to make for an automated reasoning system, unless it would have access to all relevant common background knowledge. We can of course use computational support for finding proofs or models. Notice that humans theorizers have often difficulty in finding those counterexamples that are non-intended models. Theorizers tend to ignore these models since conflict with their common-sense understanding of the domain. Fortunately, an automated model generator does not have this bias and often can reveal the counterexamples due to implicit background knowledge within seconds.

Each of these revisions takes into account a particular (set of) models or counterexamples. Note that this results in a piecemeal revision: after any revision, further testing may reveal different counterexamples that can be addressed separately. As a result, the axiomatization of informal theories proceeds in a cyclic process, in which the formal theory is repeatedly revised. More importantly, these modifications of the formal theory may have an impact on the original theory. Consider the case in which the original theory contains hiatus. If we can find reasonable assumptions that make the conjectures derivable in the formal theory, we can, again, translate these added assumptions back to the original theory. It may also be the case that the formal theory reveals that certain restricting assumptions of the original theory are not necessary or can be relaxed. In short, in these cases the formal theory and the original theory will evolve in parallel. For examples of this, the reader is referred to the axiomatizations of social science theories [including Péli et al., 1994, Hannan, 1998, Kamps and Pólos, 1999].

The *product* of an axiomatization attempt, a first order logic rendition of a theory, is a deductive theory. However, the *process* to axiomatize a theory is essentially is non-deductive, since we will typically decide to revise the theory to resolve various deficiencies. For example, the step to revise the theory to account for a counterexample, is abductive. In fact, this step is using an extended form of abduction, for want of a better name call it model-based abduction. We may either decide to strengthen the premises to explain the conjecture (as in the traditional form of abduction [Aliseda, 1997]), or decide to weaken the conjecture such that it can be explained by the original premises. Note that only in the first case we really modify the formal theory (being all logical consequences of the premise set), in the second case only the exposition of the theory changes. Our approach was to clearly distinguish between revising the theory (either by adding implicit background knowledge, or a proper revision by making the premises stronger such that the models that were counterexamples are no longer models of the theory), or revising the intended theorem (by making the conjecture weaker such that it will be true in the models that were counterexamples). The abduction literature generally does not make this distinction, and seems to cover only the case in which the theory is revised.[2]

Viewed as a form of abduction, this type of theory revision has an interesting property. In case of traditional abduction, the abductive explanation is guaranteed to entail the observation (given the background theory). Such an explanation may very well be stronger than strictly necessary. Moreover, there may exists different minimal explanations based on different parts of the theory or knowledge base. As a result of this, abductive steps tends to be far less deterministic than their deductive counterparts (although some have ingenious ways of restricting them [Inoue, 1992]). The intuitions behind "model-based" abduction constitute a dual approach the same problem of providing explanations. A revision that eliminates a set of counterexamples is a cautious revision that is not guaranteed to make a conjecture derivable (for there may exist other counterexamples). In other words, such a revision may very well be weaker than necessary. On the other hand, every one of the conceivable revisions we could be making in case of traditional abduction should address all these counterexamples—if the chosen revision

[2]However, note that an abductive explanation consists typically of the appropriate initial conditions that allow for a singular fact to be deduced from a general knowledge base. This seems intuitively a revision of the 'conjecture' rather than of the 'theory'—surely, one does not want to include particular initial conditions as part of the theory? There seems to be a fundamental difference between the revision of a theory and the reformulation of a conjecture, even if the resulting explanations are in a certain sense the same. Using the *deduction theorem*: $\Sigma \cup \{\phi\} \vdash \psi$ if and only if $\Sigma \vdash (\phi \rightarrow \psi)$ (with ϕ having no free variables [Chang and Keisler, 1990]). Whether this distinction is also relevant for other domains depends on the specific knowledge base at hand.

fails to address one of these models, this counterexample will remain, and we will still be unable to derive the conjecture.

References

A. Aliseda. *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. ILLC dissertation series 1997-4, Department of Philosophy, Stanford University, Stanford CA, 1997.

C. C. Chang and H. J. Keisler. *Model Theory*, volume 73 of *Studies in Logic and the foundations of mathematics*. North-Holland, Amsterdam, third edition, 1990.

A. Fisher. *The logic of real arguments*. Cambridge University Press, Cambridge UK, 1988.

M. T. Hannan. Rethinking age dependence in organizational mortality: Logical formalizations. *American Journal of Sociology*, 104:126–164, 1998.

K. Inoue. Linear resolution for consequence finding. *Artificial Intelligence*, 56: 301–353, 1992.

J. Kamps. Formal theory building using automated reasoning tools. In A. G. Cohn, L. K. Schubert, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98)*, pages 478–487. Morgan Kaufmann Publishers, San Francisco CA, 1998.

J. Kamps. On criteria for formal theory building: Applying logic and automated reasoning tools to the social sciences. In J. Hendler and D. Subramanian, editors, *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*, pages 285–290. AAAI Press/The MIT Press, Menlo Park CA, 1999.

J. Kamps and L. Pólos. Reducing uncertainty: A formal theory of *Organizations in Action*. *American Journal of Sociology*, 104:1776–1812, 1999.

G. Péli, J. Bruggeman, M. Masuch, and B. Ó Nualláin. A logical approach to formalizing organizational ecology. *American Sociological Review*, 59:571–593, 1994.