The Ricean Objection: An Analogue of Rice's Theorem for First-order Theories

IGOR CARBONI OLIVEIRA and WALTER CARNIELLI, *Centre for Logic, Epistemology and the History of Science (CLE), and Department of Philosophy (IFCH), State University of Campinas - UNICAMP, C.P. 6133 - 13083-970 Campinas, SP, Brazil.*

Abstract

We propose here an extension of Rice's Theorem to first-order logic, proven by totally elementary means. If P is any property defined over the collection of all first-order theories and P is *non-trivial* over the set of finitely axiomatizable theories (i.e., P holds for some, but not all theories), then P is undecidable. This not only means that the problem of deciding properties of first-order theories is as hard as the problem of deciding properties about languages accepted by Turing machines, but also offers a general setting for proving several undecidability results in first-order theories.

1 Undecidability in logic and computer science

The undecidability of computational problems by Turing machines (TM) is a formidable theoretical result that affects the practice of computer programming. Rather than being isolated, undecidability is an ubiquitous phenomenon, in the sense that any relevant (non-trivial, in the sense of holding for some pograms but not all of them) property about computer programs is undecidable: indeed, H. G. Rice proved more than half a century ago (cf. [Ri]) that any property defined over the set of languages accepted by Turing machines is either trivial or undecidable. Rice's theorem has an equivalent (perhaps better known) statement in the language of recursive function theory, and it is reasonable to ask for the effects of such "Ricean" objection to logic in general. Some previous work already point to this direction: in [GMS], a book devoted to studying computer models in philosophical research, the first chapter contains two theorems on undefinability of chaos which bear some similarity with Rice's (sometimes called Rice-Shapiro) theorem in its recursion-theoretical version. A topological version of Rice-Shapiro Theorem appears in [HM].

The uncomputability barriers of Ricean theorems poses hard challenges for the understanding and characterization of the distinction between what is known in programming jargon as "extensional" properties of programs, (i.e. those properties of the class of input-output pairs computed by a program), and the "intentional" properties of programs, (i.e. properties of the program code): indeed, the first will suffer from undecidability, while the latter are generally decidable. In logic, analogous to this difference would be the distinction between expressability (properties of theories) versus syntactic-linguistic features of logic theories.

Our main result shows that universal undecidability properties granted in computation are also shared by the first-order logical theories. The conclusion is that the problem of

Vol. 16 No. 6,
doi:10.1093/jigpal/jzn023

deciding properties of mathematical theories is as hard as that of deciding properties about the language accepted by Turing machines, and this may have interesting consequences.

We prove in this note, by an elementary argument, that these ideas can be extended to formal systems not directly related with the notion of computation. By defining for first-order theories analogous concepts as the involved in Rice's original result, we can show a general form of this theorem to be valid for first-order theories as well.

In the next section we discuss the ideas and concepts involved. The formal statement of the problem and its proof is given in the subsequent section.

2 Statement of the problem

Given a Turing machine M, let $Lang(M)$ be the set of words accepted by M. Some immediate questions about the language represented by the TM M can be posed:

- Does the empty word Λ belongs to $Lang(M)$?
- Is $Lang(M)=\Sigma^*$, i.e., is it the case that every word over the input alphabet Σ of the TM M is accepted by M?

It is possible to show that these and several other problems of the same kind are algorithmically unsolvable. A starting technique for such proofs is to show, by diagonalization, that a first problem is undecidable and then prove, using reduction, that so are the other problems. It is usual to use, as first step, the problem of deciding whether a TM M accepts a word w of its input alphabet:

$$A_{TM}=\{\langle M,w\rangle \mid M \text{ is a TM and } M \text{ accepts the word } w\}.$$

In the previous notation, $\langle M,w\rangle$ is a reasonable codification of an instance of the problem. A. Turing [Tu] proved that A_{TM}, known as the halting problem, is undecidable by Turing machines. The undecidability of the halting problem and some other problems are special cases of the next result.

Rice's Theorem. Let P be a property of languages that is satisfied by some, but not by all, recursively enumerable languages. Then the following problem is undecidable:

D_P: Given a TM M, does $Lang(M)$ satisfy P?

The proof of this remarkable result can be found in Rice [Ri].

We can interpret a TM M as a finite representation of the possibly infinite set $Lang(M)$. In exactly the same way, if we interpret a logical theory T as a set of formulas from a specific first-order language, we can try to represent it by a finite set A of axioms (formulas). If $Th(A)=T$, i.e., the set of formulas provable from A is the set T, then we have found a suitable representation for the theory T. It is possible to proceed in a similar way and formulate some questions about the theory T, finitely represented by A.

- Is T a complete theory?
- If ψ is a formula in the language of T, does $\psi \in T$?
- Is T a decidable theory?

To answer this and other questions, we chose a sufficiently expressive first-order language. We assume that the language of all the first-order theories refereed will be that generated by the signature $\Sigma = \{0, s, +, *\}$, where 0 is a constant, s is an unary function and $+$ and $*$ are binary functions. Denote by L_Σ the set of valid formulas generated by this signature. This is not a great restriction since the result we present can be easily translated to richer first-order languages.

If a theory T is finitely axiomatizable, its codification will be denoted by $\langle T \rangle$. In this way, we are only interested in properties of theories that are finitely axiomatizable, and hence finitely representable.

The next results show that some properties of first-order theories are undecidable. Let Q be the finite set of axioms for weak Arithmetic as presented in [EC].

Theorem 1. *$Th(Q)$ is an essentially undecidable theory.*

Proof. See [EC]. ∎

Theorem 2. *Let $A_{LT} = \{\langle T, \psi \rangle \mid$ the formula ψ is provable in the theory $T\}$. Then A_{LT} is undecidable.*

Proof. As Q is a particular instance of A_{LT}, the result readily follows from theorem 1. ∎

Using the reduction technique we will prove some other properties to be undecidable. Consider the following set:

$CONSIS = \{\langle T \rangle \mid T$ is a consistent theory $\}$.

Theorem 3. CONSIS *is undecidable by Turing machines.*

Proof. Suppose that *CONSIS* is decidable by TM M_1. We define the following TM M_2 to decide A_{LT}:

If $\langle T, \psi \rangle$ is the input of M_2, then:

1- M_2 creates the theory $T' = T \cup \{\neg \psi\}$, codified by $\langle T' \rangle$.
2- M_2 asks whether TM M_1 accepts $\langle T' \rangle$.
3- If M_1 accepts $\langle T' \rangle$, then M_2 rejects $\langle T, \psi \rangle$, otherwise M_2 accepts $\langle T, \psi \rangle$.

Suppose that ψ is a thorem of T. By its definition, M_2 creates the theory $T' = T \cup \{\neg \psi\}$ and sends it to the TM M_1. As M_1 decides *CONSISTENT* and the theory T' é inconsistent, M_1 rejects T' and because of this M_2 accepts $\langle T, \psi \rangle$. Now suppose that ψ is not a theorem of T. Then $T' = T \cup \{\neg \psi\}$ is consistent and is accepted by M_1. Hence M_2 rejects $\langle T, \psi \rangle$. That is, M_2 accepts $\langle T, \psi \rangle$ if and only if ψ is a theorem of T, i.e., M_2 decides A_{LT}. By theorem 2, this is impossible. We conclude that the Turing machine M_1 does not exist and hence *CONSIS* is undecidable. ∎

For now on, the detailed steps in the construction of the TM will be omitted. The next result can also be found in A. Tarski [Ta]. Consider the following set:

$DECIDABLE = \{\langle T \rangle \mid T$ is a decidable theory $\}$.

Theorem 4. DECIDABLE *is a set that is undecidable by Turing machines.*

Proof. For the sake of a contradiction, assume that this set is decidable. As a particular case of a first-order theory, we have the undecidable theory Q. By theorem 1, every finite consistent extension of Q is also undecidable. Hence, if $Q \cup \{\neg\psi\}$ is consistent, then $Q \cup \{\neg\psi\}$ is undecidable. On the other hand, if $Q \cup \{\neg\psi\}$ is inconsistent, then $Q \cup \{\neg\psi\}$ is decidable. But we know that the formula ψ is not a theorem of Q if and only if $Q \cup \{\neg\psi\}$ is consistent. Therefore:

$Q \vdash \psi$ *iff* $Q \cup \{\neg\psi\}$ is inconsistent *iff* $Q \cup \{\neg\psi\}$ is decidable.

By theorem 1, we conclude that *DECIDABLE* is undecidable by Turing machines. ∎

Following the same ideas related to the original result of Rice, we propose the following problem:

Problem. *Let P be a property defined over the subsets of formulas of L_Σ. Given a theory T, is it the case that $Th(T)$ satisfies P?*

In the next section we give necessary and sufficient conditions for the property P to be undecidable.

3 A Rice theorem for logic

This section formalizes the previously stated problem and prove a version of Rice's theorem for first-order theories.

Definition. *Let $\Sigma = \{0, s, +, *\}$ be a signature where 0 is a constant, s is an unary function and $+$ and $*$ are binary functions. Let L_Σ be the set of valid formulas generated by the signature Σ.*

Definition. *A theory T is a subset of L_Σ. If $A \subset L_\Sigma$, then $Th(A)$ is the set of formulas (theorems) of L_Σ derivable from A. T is a finitely axiomatizable theory if $T = Th(A)$ for some finite set A of formulas.*

Definition. *A property is any subset Prop of the power set $Pow(L_\Sigma)$ of the set of formulas of L_Σ.*

Definition. *Let T_{Fin} be the set of finitely axiomatizable theories. A property Prop is trivial if it is either satisfiable by every element of T_{Fin} or if it is not satisfied by any element of T_{Fin}.*

Definition. *Prop is a decidable property if, given any finite set A of formulas in the language L_Σ, there exists an algorithm that decides if $Th(A) \in Prop$. Prop is a finite property if Prop is a finite set.*

The following lemma will be useful.

Lemma 1. *Let T be a finitely axiomatizable and undecidable theory. Let I be the set of formulas in the language of T that are undecidable by T (i.e., $\phi \in I \Leftrightarrow T \nvdash \phi$ and $T \nvdash \neg\phi$). Then the set I is undecidable.*

Proof. Suppose I is decidable. Then the following algorithm decides T: given ϕ, verify if ϕ belongs to I. In the positive case, then $T \nvdash \phi$. On the other hand, then $T \vdash \phi$ or $T \vdash \neg\phi$. In the last case, all we need to do is generate the theorems of T until one of the two formulas is found, and then give the correct answer. It is easy to see that this procedure always halts and that it is able to determine if a formula is a theorem of T. ∎

Rice's theorem for logic. *If Prop is a nontrivial property of first-order theories, then Prop is undecidable.*

Proof. Let Q be the same theory defined previously. Consider, without loss of generality, that $L_\Sigma \in Prop$. As *Prop* is nontrivial, there exists a finitely axiomatizable theory T such that $T \notin Prop$. Consider the (finitely axiomatizable) theory U composed by the following axioms:

Ax1: $(Q \Rightarrow \phi) \Rightarrow \psi \wedge \neg\psi$
Ax2: $(Q \Rightarrow \neg\phi) \Rightarrow \psi \wedge \neg\psi$
Ax3: $[\neg(Q \Rightarrow \phi) \wedge \neg(Q \Rightarrow \neg\phi)] \Rightarrow T$

where ϕ, ψ are arbitrarily formulas and Q and T represent the formulas that are the conjunction of the axioms of Q and T, respectively. Suppose that ϕ is undecidable in Q, i.e., $Q \nvdash \phi$ and $Q \nvdash \neg\phi$. In this case, $Ax1$ and $Ax2$ are tautologies, and therefore $Th(U) = Th(T)$. Hence $U \notin Prop$. On the other hand, if ϕ is decidable in Q, it is easy to see that $U = L_\Sigma$ and therefore $U \in Prop$. Hence, ϕ is undecidable in Q if and only if $U \notin Prop$. If *Prop* is decidable, then the set of undecidable formulas of Q is also decidable. But this contradicts Lemma 1. Therefore, *Prop* is undecidable. ∎

A simple application of this result proves, for instance, that there is no algorithm that is able to decide if two set of axioms give rise to the same theory.

4 Conclusion

Several (if not most) mathematical theories are undecidable. These so-called classical decision problems have been intensively investigated (see, for instance, [BGG]), and it is well accepted today that proofs of undecidability basically rely on extensions of Gödel's well-known incompleteness arguments, or on extensions of the reduction method originated by Tarski, which of reducing a theory under question to a previously known undecidable theory (see e.g. [Mu] for a condensed historical and conceptual account of the adventure of decidability and undecidability). The pioneering method for proving the undecidability of the *Entscheidungsproblem* of first-order logic by A. Church in 1936 can be seen as a common precursor of the two main methods, and somehow connects undecidability results in computability and logic. Our main result shows that universal undecidability properties behind non-computability are also shared by the first-order logical theories. A consequence is that the problem of deciding properties of mathematical theories is as hard as that of deciding properties about languages accepted by Turing machines.

Acknowledgements

This research was supported by FAPESP- Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, Thematic Research Project grant 2004/14107-2. The first author is also supported by a FAPESP Scientific Initiation grant, and the second by a CNPq (Brazil)

Research Grant 300702/2005-1 and by the Fonds National de la Recherche Luxembourg. We thank R. L. Epstein for criticisms and comments.

References

[BGG] E. Börger, E. Grädel and Y. Gurevich. The Classical Decision Problem. Springer Verlag, Berlin, 1997.

[EC] R. L. Epstein and W. A. Carnielli. Computability, Computable Functions, Logic, and the Foundations of Mathematics. 3rd Edition. Advanced Reasoning Forum, 2008.

[GMS] P. Grim, G. Mar and P. St. Denis. The Philosophical Computer. Exploratory Essays in Philosophical Computer Modeling. The MIT Press, Cambridge, Mass., and London, 1998.

[HM] L. Hay and D. Miller. A topological analog to the Rice-Shapiro index theorem. *The Journal of Symbolic Logic* 47: 4 1982, 824–832.

[Mu] R. Murawski. Decidability vs. undecidability. Logico-philosophico-historical remarks. *Annales UMCS Informatica AI* 3 (2005), 105–117.

[Ri] H. G. Rice. Classes of Recursively Enumerable Sets and Their Decision Problems. *Trans. Amer. Math. Soc.* 74, 358–366, 1953.

[Ta] A. Tarski. Undecidable Theories. North-Holland, 1953.

[Tu] A. M. Turing. On Computable Numbers, with an Application to the Entsheidungsproblem. *Proceedings of the London Mathematical Society*, Series 2, 42 (1936–37), pp. 230–265.