Jouko Väänänen
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
The Netherlands
vaananen@science.uva.nl

How complicated can structures be?

Is there a measure of how 'close' non-isomorphic mathematical structures are? Jouko Väänänen, professor of logic at the Universities of Amsterdam and Helsinki, shows how contemporary logic, in particular set theory and model theory, provides a vehicle for a meaningful discussion of this question. As the journey proceeds, we accelerate to higher and higher cardinalities, so fasten your seatbelts.

By a *structure* we mean a set endowed with a finite number of relations, functions and constants. Examples of structures are groups, fields, ordered sets and graphs. Such structures can have great complexity and indeed this is a good reason to concentrate on the less complicated ones and to try to make some sense of them. In this article we walk the less obvious and perhaps less appealing trail of delving more deeply into more and more complicated structures. We raise the question of how we can make sense of the statement that we have found an extremely complicated structure? This is typical of the kind of question investigated in mathematical logic. The guiding result of mathematical logic is the Incompleteness Theorem of Gödel, which says that the logical structure of number theory is so complicated that it cannot be effectively axiomatized in its entirety. In other words, the theory is non-recursive, i.e. there is no Turing machine that could tell whether a sentence of number theory is true or not. A contrasting and pivotal result of logic from the same period is Alfred Tarski's result that the field of real numbers (or the field of complex numbers) can be completely and effectively axiomatized and is indeed recursive in the sense that there is a Turing machine that decides whether a given statement about the plus and times of real (or complex) numbers is true or not.

We start with the extremely interesting situation concerning attempts to classify finite models. We then move to the more established case of countable structures. Sweeping results exist here and this case is very much the focus of current research. Then we turn our faces to the wind and stare into the eyes of the difficult uncountable structures. New ideas are needed here and a lot of work lies ahead. Finally we tie the uncountable case to stability theory, a recent trend in model theory. It turns out that stability theory and the topological approach proposed here give similar suggestions as to what is complicated and what is not.

For unexplained set theoretical concepts refer to [5].

Finite structures

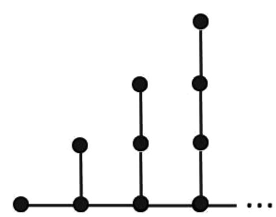
Let us start with finite structures. The famous P=NP question, one of the Clay Institute Millennium Questions, asks if we can decide in polynomial time whether a given finite graph is 3-colourable. Should the answer to the P=NP question be negative, as is expected, we will have a sequence of some rather complicated graphs, for which no algorithm, running in polynomial time in the size of the graph, can decide whether the graph is 3-colourable or not.

The problem of whether the isomorphism of two finite structures can be solved in polynomial time is a famous open problem of complexity theory. It is particularly famous because it is not known whether it is NP-complete either; it may be strictly between P and NP.

Countable structures

What about countably infinite structures? We should not distinguish between isomorphic structures. So let us assume the universe of our countable structures is the set $\mathbb{N}$ of natural numbers. After a little bit of coding, such countable structures can be thought of as points in the topological space N of all functions $f : \mathbb{N} \to \mathbb{N}$ endowed with the topology of pointwise convergence, where $\mathbb{N}$ is given the discrete topology.

We can now consider the orbit of an arbitrary countable structure under all permutations of $\mathbb{N}$ and ask how complex this set is in the topological space N. If the orbit is a closed set in this topology, we should think of the structure as an uncomplicated one. This is because the orbit being closed means, in view of the definition of the topology, that the finite parts of the structure completely determine the whole structure, as is easily seen to be the case in the graph of the picture:



A structure may be quite innocuous even if the orbit is not closed. For example, the orbit of the ordered set of the rationals is not closed because as far as the finite parts are concerned it cannot be distinguished from the order type of the integers. While not closed, the orbit of the rationals is of the form

$$\bigcap_n \bigcup_m F_{n,m}, \qquad (1)$$

where each $F_{n,m}$ is closed. This is a consequence of the fact that the density of the order, as well as not having endpoints, can be expressed in the form 'for all ... exists ...', and these two properties completely determine the structure among countable structures.

When the number of alternating intersections and unions increases in the formula (1), even to the transfinite, we end up with the hierarchy of Borel sets

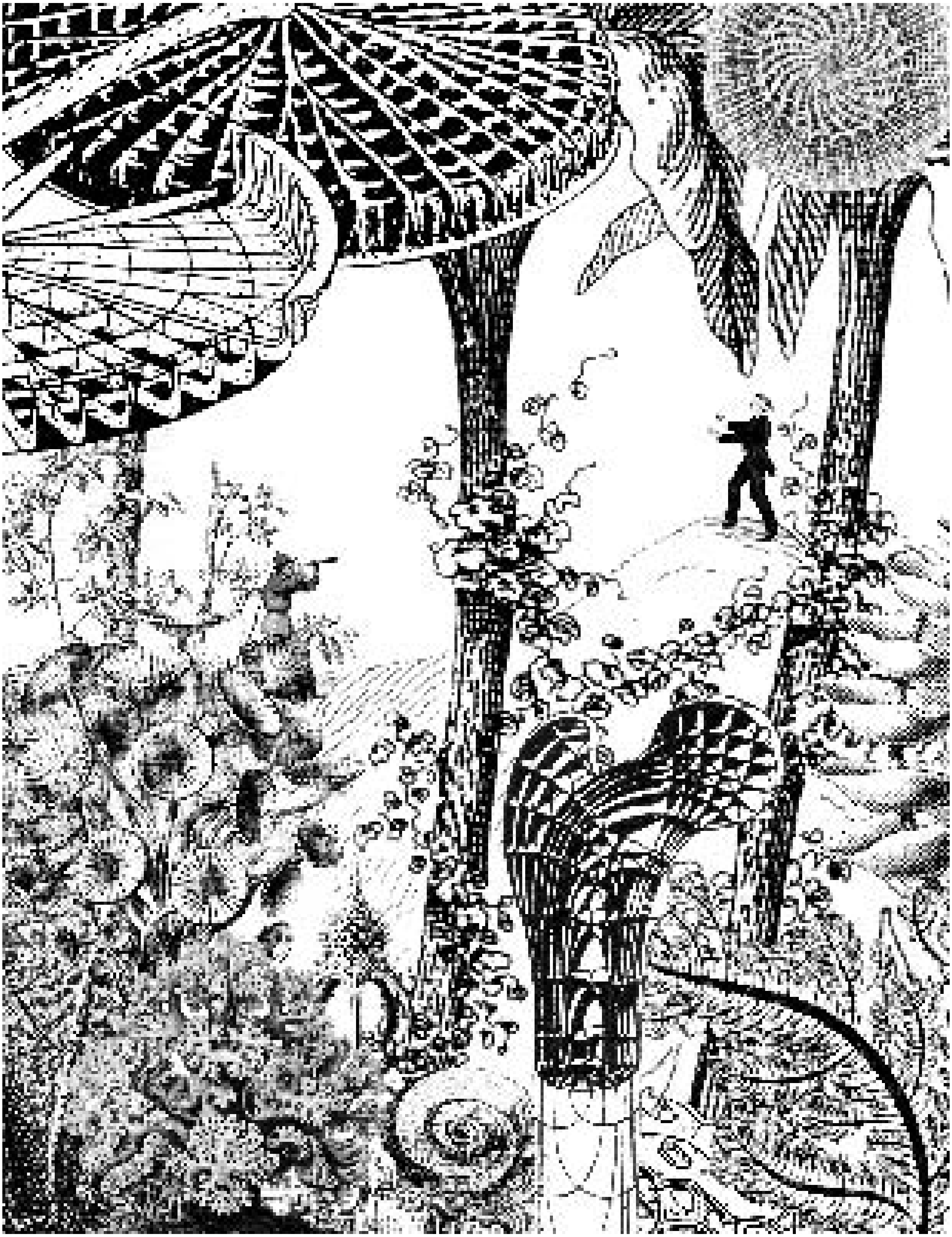


Illustration: Ryu Tajiri

- G_0 = open sets, F_0 = closed sets
- $G_{\alpha+1}$ = countable unions of sets from F_α
- $F_{\alpha+1}$ = countable intersections of sets from G_α
- $G_\nu = \bigcup_{\alpha<\nu} G_\alpha$, $F_\nu = \bigcup_{\alpha<\nu} F_\alpha$, if ν is a limit ordinal

named after Émile Borel (1871–1956), a French mathematician.

So the philosophy is now that the further the orbit is from being a closed set the more complicated the structure is. We can go up the Borel hierarchy and find structures on all levels $F_\alpha \cup G_\alpha$. By a deep result of Dana Scott [9] every orbit is on some level of the Borel hierarchy, although *a priori* the orbits are just analytic sets, i.e. continuous images of closed sets.

The levels $F_\alpha \cup G_\alpha$ of the Borel hierarchy are calibrated by countable ordinals α. Orbits of familiar structures such as $(\mathbb{N}, +, \cdot, 0, 1)$, the field of rational numbers, the Random Graph, the free Abelian group of countably many generators, and any vector space (over $\mathbb{Q}$) of countable dimension are all on one of the lowest infinite levels of the Borel hierarchy. On the other hand the orbit of any sufficiently closed countable ordinal $(\alpha, <)$ is on level α, i.e. in the set F_α but not in any $F_\beta \cup G_\beta$ for $\beta < \alpha$ (what is needed is that α is such that $\beta < \alpha$ implies $\omega^\beta < \alpha$). Such structures of high level can be constructed for e.g. Abelian groups. This basic setup has led recently to a rich theory of Borel equivalence relations on Polish spaces [1].

Uncountable structures

What if we have an uncountable structure and we want to measure its complexity and the degree to which a given structure is close to being isomorphic to it? After all, the most important mathematical structures, such as the fields of real numbers and complex numbers, Euclidean spaces, Banach spaces, etc., are all uncountable. In the light of our experience with countable structures, it seems natural to consider structures that are determined by their countable parts as uncomplicated.

Consider the ordered set $L = (\mathbb{R}, <)$ of all real numbers. If we only look at countable sub-orders, this is no different from the ordered set $L' = (\mathbb{R} \setminus \{0\}, <)$ of the non-zero real numbers — although L and L' are not isomorphic as the first is a complete order and the second is not. In fact, L is quite a complicated structure albeit not by any means among the most complicated. One example of the peculiar properties of L is the following. If we add a new real to the universe by Cohen's method of forcing, L becomes isomorphic with L'. So in some sense L is a hair's breadth away from being L'. The fact that L is complicated is related to exactly this kind of phenomenon, to being an iota away from another, non-isomorphic structure.

When we set our foot on the path of looking at uncountable structures through the lens of their countable parts, the first resting spot is bound to be the class of structures that can be expressed as an increasing union of countable substructures or, equivalently, structures of cardinality $\aleph_1$, the first uncountable cardinal. Now the alarm bells start to ring! We do not know whether the real numbers, the complex numbers, Euclidean space, Banach spaces, etc, have this property.

The question of whether the set $\mathbb{R}$ of real numbers is an increasing union of countable sets is known as the Continuum Hypothesis (CH). So in order to include those structures in this discussion we have to assume CH. In fact, most of the currently known results in this direction assume CH anyway. But there is a whole family of structures that are by their very definition increasing unions of countable structures, and this family is closed under various algebraic operations but not under infinite products, unless we assume CH. An example is the order-type $(\omega_1, <)$ and the numerous structures built around it, such as the free Abelian group on $\aleph_1$ generators.

Models of cardinality $\aleph_1$ can be thought of as points in the space N_1 of functions $f : \omega_1 \to \omega_1$ endowed with the topology of pointwise convergence, that is, a neighbourhood of a point $f \in N_1$ is of the form $N(f, X) = \{g \in N_1 : \forall x \in X(g(x) = f(x))\}$ where X is countable. So the orbit of a structure is a closed set essentially if the countable parts of the structure completely determine it. The orbit of the free abelian groep $F(\aleph_1)$ on $\aleph_1$ is not at all closed. There are so-called almost free Abelian groups, every countable subgroup of which is free but which are not free themselves. So the question to ask is not only what the countable substructures are but also how they sit inside the structure. To see how complicated the group $F(\aleph_1)$ is let us define the Borel hierarchy in N_1.

The class of *Borel* sets of the space N_1 is the smallest class of sets containing the open sets and closed under complements and unions of length ω_1. A set is *analytic* if it is a continuous image of a closed subset of N_1. Orbits of structures of cardinality $\aleph_1$ are, *a priori*, analytic, but are they Borel?

When we carry out the same topological analysis of models of cardinality $\aleph_1$, as we did with countable models, the notion of an approximation is more complex. After all, we have to approximate an uncountable object and we cannot approximate $f \in N_1$ merely by its finite initial segments. Roughly for the same reason, the approximations are scaled by bounded trees, i.e. trees with no uncountable branches, rather than ordinals.

The passage from well-founded trees to bounded trees brings with it two major problems. The first is the problem of the ordering of the class of all such trees. The problem of the ordering of the trees is the following. In analogy with ordinals and well-founded trees we write $T \leq T'$ if there is a strict tree-order preserving (but not necessarily one-to-one) mapping from T to T'. For well-founded trees this quasi-order is connected, i.e. any two well-founded trees are comparable by $\leq$. For non-well-founded trees this need not be the case [4, 13]. This is on the one hand a shortcoming of the whole approach, as it means that some structures are incomparable as to their complexity. On the other hand it has un-

Émile Borel (1871–1956), a French mathematician and politician who has many theorems named after him; there is even a Borel crater on the moon in the *Mare Serenitatis*. Borel, together with Lebesgue and Baire, is also known as a representative of *semi-intuitionism*, an alternative approach to constructivism along with Brouwer's *intuitionism*. The former maintained that set theory should be limited to definable sets, anticipating *descriptive set theory*, while the latter launched a criticism of the Law of Excluded Middle, leading to modern *intuitionistic logic* and *constructive mathematics*.

earthed a rich theory of trees, and whatever progress we can make in this direction is directly reflected in our ability to measure how close uncountable structures can be to each other in the given topology. One puzzle that has arisen in this connection is the existence of a *Canary tree*. This name is due to the following special role of Canary trees. If any stationary subset of ω_1 is killed by forcing without adding reals, then the Canary tree gets a long branch. Summing up, if a stationary set is poisoned somewhere then the Canary tree warns us by expiring. The existence of Canary trees cannot be decided on the basis of ZFC or even CH alone [7]. However, Canary trees are intimately related to the complexity of some canonical structures. It can be shown that there is a Canary tree if and only if the orbit of the free Abelian group of $\aleph_1$ generators is analytic co-analytic [6]. In the space N, analytic co-analytic sets are Borel but in the space N_1, the situation is more complicated. The most promising attempt to bring order into the chaos of bounded trees is the approach of Todorcevic [12] under the assumption of the so-called Proper Forcing Axiom (PFA). This axiom says, very roughly speaking, that the universe is invariant under changes imposed by a certain restricted form of Cohen's concept of forcing. Another approach is to restrict to sufficiently definable trees and thereby avoid the incomparability problem [2].

Another new feature that arises in the study of N_1 is the fact that if we assume CH, the Luzin Separation Principle fails. There are disjoint analytic sets that cannot be separated by a Borel set [11]. This further emphasizes how the difference between the countable and the uncountable is reflected in the topology of N and N_1, and thereby in the classification of countable versus uncountable models.

Stability theory

In modern model theory there is an alternative approach to the problem of classifying structures, namely stability theory [8]. The difference is that stability theory tries to classify complete first order theories rather than structures. However, the message of stability theory is that all models of size $\aleph_1$ of theories satisfying a combination of certain stability conditions (superstable, NDOP, DOTOP) are rather 'uncomplicated' in the sense that their isomorphism can be expressed in terms of a determined game (called the Ehrenfeucht-Fraïssé-game) of length ω [10]. Naturally, such structures may be very complicated in other ways. The point of stability theory is that in such structures one can define a kind of geometry that enables one to classify the structure in terms of dimension-like invariants. On the other hand, theories failing to satisfy such stability conditions are bound to have models of cardinality $\aleph_1$ that are extremely complicated. This is Shelah's 'Main Gap' [10]. For example, assuming CH and if there are no Canary trees, such theories have models of cardinality $\aleph_1$ with high complexity in the definability theoretic sense described in this article [3]. To measure the height of the complexity one uses bounded trees and it turns out that under the stated assumptions one can go beyond any bounded tree. Work in this direction is very much underway.

Conclusion

The study of the complexity of uncountable structures is an interdisciplinary subject. We need to develop set theory, and especially the theory of trees, in order to have a good measure of the complexity of uncountable models. At the same time, we have to develop model theory, and especially stability theory, in order to distinguish important dividing lines between simple and complicated structures. Both set theory and model theory suggest that we should look for an answer in the direction of long games. In model theory the relevant games are known as Ehrenfeucht-Fraïssé games. In set theory the corresponding games are related to the so-called stationary sets. The results referred to above connect the two games and thereby tie a knot connecting set theory and model theory. When we look into the deep eyes of the uncountable structures, we are perhaps starting to see there some compassion for our modest advances, our budding infinite trees, our courageous appeals to stability and our resolve to play the game to the end.

References

1 Howard Becker and Alexander S. Kechris, *The descriptive set theory of Polish group actions*, volume 232 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

2 Ilijas Farah and Paul B. Larson, 'Absoluteness for universally Baire sets and the uncountable' I. In *Set theory: recent trends and applications*, volume 17 of *Quad. Mat.*, pages 47–92. Dept. Math., Seconda Univ. Napoli, Caserta, 2006.

3 Tapani Hyttinen and Heikki Tuuri, 'Constructing strongly equivalent nonisomorphic models for unstable theories', *Ann. Pure Appl. Logic*, 52(3):203–248, 1991.

4 Tapani Hyttinen and Jouko Väänänen, 'On Scott and Karp trees of uncountable models', *J. Symbolic Logic*, 55(3):897–908, 1990.

5 Kenneth Kunen, *Set theory*, volume 102 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, 1983. An introduction to independence proofs. Reprint of the 1980 original.

6 Alan Mekler and Jouko Väänänen, 'Trees and Π^1_1-subsets of ${}^{\omega_1}\omega_1$', *J. Symbolic Logic*, 58(3):1052–1070, 1993.

7 Alan H. Mekler and Saharon Shelah, 'The canary tree', *Canad. Math. Bull.*, 36(2):209–215, 1993.

8 Anand Pillay, *Geometric stability theory*, volume 32 of *Oxford Logic Guides*, The Clarendon Press Oxford University Press, New York, 1996. Oxford Science Publications.

9 Dana Scott, 'Logic with denumerably long formulas and finite strings of quantifiers', In *Theory of Models (Proc. 1963 Internat. Sympos. Berkeley)*, pages 329–341. North-Holland, Amsterdam, 1965.

10 S. Shelah, *Classification theory and the number of nonisomorphic models*, volume 92 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, second edition, 1990.

11 Saharon Shelah and Jouko Väänänen, 'Stationary sets and infinitary logic', *J. Symbolic Logic*, 65(3):1311–1320, 2000.

12 Stevo Todorčević, 'Lipschitz maps on trees', *J. Inst. Math. Jussieu*, 6(3):527–556, 2007.

13 Stevo Todorčević and Jouko Väänänen, 'Trees and Ehrenfeucht-Fraïssé games', *Ann. Pure Appl. Logic*, 100(1-3):69–97, 1999.