On Ordinal VC-Dimension and Some Notions of Complexity

Eric Martin[1], Arun Sharma[2], and Frank Stephan[2]

[1] School of Computer Science and Engineering, UNSW Sydney, NSW 2052, Australia
emartin@cse.unsw.edu.au
[2] National ICT Australia, UNSW Sydney, NSW 2052, Australia
{Arun.Sharma, Frank.Stephan}@nicta.com.au

Abstract. We generalize the classical notion of VC-dimension to *ordinal VC-dimension*, in the context of *logical* learning paradigms. Logical learning paradigms encompass the numerical learning paradigms commonly studied in Inductive inference. A logical learning paradigm is defined as a set $\mathcal{W}$ of structures over some vocabulary, and a set $\mathcal{D}$ of first-order formulas that represent data. The sets of models of φ in $\mathcal{W}$, where φ varies over $\mathcal{D}$, generate a natural topology $\mathbb{W}$ over $\mathcal{W}$.
We show that if $\mathcal{D}$ is closed under boolean operators, then the notion of ordinal VC-dimension offers a perfect characterization for the problem of predicting the truth of the members of $\mathcal{D}$ in a member of $\mathcal{W}$, with an ordinal bound on the number of mistakes. This shows that the notion of VC-dimension has a natural interpretation in Inductive Inference, when cast into a logical setting. We also study the relationships between predictive complexity, selective complexity—a variation on predictive complexity—and mind change complexity. The assumptions that $\mathcal{D}$ is closed under boolean operators and that $\mathbb{W}$ is compact often play a crucial role to establish connections between these concepts.

Keywords: Inductive inference, logical paradigms, VC-dimension, predictive complexity, mind change bounds.

1 Introduction

The notion of VC-dimension is a key concept in PAC-learning [6, 12, 13]. The notion of finite telltale is a key concept in Inductive inference [3, 8]. It can be claimed that VC-dimension is to PAC-learning what finite telltales are to Inductive inference. Both provide a characterization of learnability, for fundamental classes of learning paradigms, in the respective settings. Both take the form of a condition where finiteness is a key requirement, in frameworks that deal with infinite objects. In logical learning paradigms of identification in the limit, it has been shown that the finite telltale condition can be seen as a generalization of the compactness property, the latter being the hallmark of, equivalently, finite learning or deductive inference [10]. The finite telltale condition can even be generalized and be interpreted as a property of *β-weak compactness*, that characterizes classification with less than β mind changes [10]. There are extensions of

VC dimension to infinite domains [4]. But there are few essential connection between VC-dimension and some fundamental concepts from Inductive inference: the relevance of the concept of VC-dimension seems to be closely related to the existence of probability distributions over the sample space. Though connections exists between PAC learning and Inductive inference (*e.g.*, [5]), it does not seem that VC-dimension has any chance to play a key role in learning paradigms of Inductive inference that do not introduce probability distributions over the sample space. We will show that VC-dimension can still provide a perfect characterization for the problem of predicting whether a possible datum is true or false in the underlying world, in the realm of Inductive inference. But for such a characterization to be possible, the condition that the set of possible data is closed under boolean operators has to be imposed. The fact that we work in a logical setting is of course essential to express this condition in a simple, meaningful and natural way. The notions and main results are stated with no computability condition on the procedures that analyze the data and output hypotheses. This is necessary to obtain perfect equivalences, and provides strong evidence that the concepts involved are naturally connected. When computability is a requirement, the relationships become more complex. Our aim is to encourage the study of the connections between ordinal VC-dimension and predictive complexity in paradigms of Inductive inference. Obtaining perfect connections for ideal, unconstrained paradigms, suggests that further work in the same direction should be carried out, in the context of more realistic or constrained paradigms. Moreover, the results will be illustrated with examples that always involve effective procedures. We proceed as follows. We introduce some background notions and notation in Section 2. We define the various complexity measures in Section 3. We study the relationships between these complexity measures in Section 4. We conclude in Section 5.

2 Background

.

The class of ordinals is denoted by Ord. Let a set X be given. The set of finite sequences of members of X, including the empty sequence $()$, is represented by $X^\star$. Given a $\sigma \in X^\star$, the set of members of X that occur in σ is denoted by $\mathrm{rng}(\sigma)$. Given a finite or an infinite sequence σ of members of X and a natural number i that, is case σ is finite, is at most equal to the length of σ, we represent by $\sigma_{|i}$ the initial segment of σ of length i. Concatenation between sequences is represented by $\star$, and sequences consisting of a unique element are often identified with that element. We write $\subset$ (respect., $\subseteq$) for strict (respect., nonstrict) inclusion between sets, as well as for the notion of a finite sequence being a strict (respect., nonstrict) initial segment of another sequence. We also use the notation $\supset$. Let two sets X, Y and a partial function f from X into Y be given. Given $x \in X$, we write $f(x) =\downarrow$ when $f(x)$ is defined, and $f(x) =\uparrow$ otherwise. Given two members x, x' of X, we write $f(x) = f(x')$ when both $f(x)$ and $f(x')$ are defined and equal; we write $f(x) \neq f(x')$ otherwise. Let

R be a binary relation over a set X. Recall that R is a well-founded iff every nonempty subset Y of X contains an element x such that no member y of Y satisfies $R(y, x)$. Suppose that R is well founded. We then denote by ρ_R the unique function from X into Ord such that for all $x \in X$:

$$\rho_R(x) = \sup\{\rho_R(y) + 1 : y \in X,\ R(y, x)\}.$$

The length of R, written as $|R|$, is the least ordinal not in the range of ρ_R. Note that $|R| = 0$ iff $X = \emptyset$. For example, Figure 1 depicts a finite binary relation R of length 5. In this diagram, an arrow joins a point x to a point y iff $R(x, y)$ holds. For all points x in the field of R, the value of $\rho_R(x)$ is indicated.

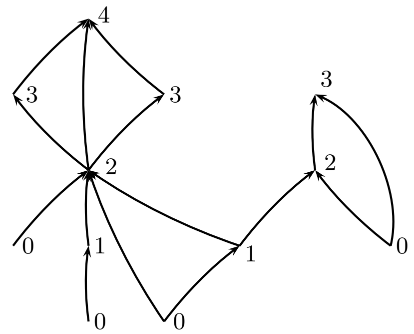

Fig. 1. A finite binary relation of length 5

Let us introduce the logical learning paradigms and their constituents. We denote by $\mathcal{V}$ a countable vocabulary, *i.e.*, a countable set of function symbols (possibly including constants) and predicate symbols. Let us adopt a convention. If we say that $\mathcal{V}$ contains $\overline{0}$ and s, then $\overline{0}$ denotes a constant and s a unary function symbol. Moreover, given a nonnull $n \in \mathbb{N}$, $\overline{n}$ is used as an abbreviation for the term obtained from $\overline{0}$ by n applications of s to $\overline{0}$.

We denote by $\mathcal{D}$ a nonempty set of first-order sentences (closed formulas) over $\mathcal{V}$ that represent data. Three important cases are sets of closed atoms (to model learning from texts), sets of closed literals (*i.e.*, closed atoms and their negations, to model learning from informants), and sets of sentences closed under boolean operators, a natural example being the set of quantifier free sentences. Note that quantifier free sentences convey no more information than closed literals. Still, the assumption that $\mathcal{D}$ is closed under boolean operators will play a key role in this paper. We denote by $\sharp$ some symbol to be used when no datum is presented. Given a member σ of $(\mathcal{D} \cup \{\sharp\})^\star$, we set $\mathrm{cnt}(\sigma) = \mathrm{rng}(\sigma) \cap \mathcal{D}$.

We denote by $\mathcal{W}$ a nonempty set of structures over $\mathcal{V}$. An important case is given by the set of all Herbrand structures, *i.e.*, structures over $\mathcal{V}$ each of whose individuals interprets a unique closed term. (When we consider Herbrand structures, $\mathcal{V}$ contains at least one constant). Given a member $\mathfrak{M}$ of $\mathcal{W}$ and a set E of formulas, the E-diagram of $\mathfrak{M}$, denoted $\mathrm{Diag}_E(\mathfrak{M})$, is the set of all members of E that are true in $\mathfrak{M}$. We say that a set T of first-order formulas is *consistent in* $\mathcal{W}$ iff T has a model in $\mathcal{W}$. Given a set T of first-order formulas over $\mathcal{V}$, we denote by $\mathrm{Mod}(T)$ the set of models of T, and by $\mathrm{Mod}_{\mathcal{W}}(T)$ the set of models of T in $\mathcal{W}$ (*i.e.*, $\mathrm{Mod}_{\mathcal{W}}(T) = \mathrm{Mod}(T) \cap \mathcal{W}$).

We denote by $\mathcal{P}$ the triple $(\mathcal{V}, \mathcal{D}, \mathcal{W})$. We call $\mathcal{P}$ a *logical paradigm*. This is a simplification of the notion of logical paradigm investigated in [9, 10]. Learning paradigms in the numerical setting are naturally cast into the logical setting as follows. Set $\mathcal{V} = \{\overline{0}, s, P\}$ where P is a unary predicate symbol. Let E be the set $\{P(\overline{n}) : n \in \mathbb{N}\}$. If $\mathcal{C}$ is the set of languages to be learnt, we define $\mathcal{W}$ as the set of Herbrand structures whose E-diagrams are $\{P(\overline{n}) : n \in L\}$ where L varies over $\mathcal{C}$. The choice of $\mathcal{D}$ depends on the type of data: $\mathcal{D} = E$ when data are positive, $\mathcal{D} = E \cup \{\neg\psi : \psi \in E\}$ when data are positive or negative.

3 The notions of complexity

We now define the various concepts of complexity we need in this paper, starting with VC-dimension. The notion of a set of hypothesis shattering a set of data takes the following form when hypotheses are represented as structures and data as formulas.

Definition 1. Let a set E of formulas be given. We say that $\mathcal{W}$ *shatters* E iff E is finite and for all subsets D of E, $\mathrm{Diag}_E(\mathfrak{M}) = D$ for some $\mathfrak{M} \in \mathcal{W}$.

Traditionally, the VC-dimension is defined as the greatest number n such that some set consisting of n distinct elements is shattered [13]. When such an n does not exist, the VC-dimension is considered to be either undefined or infinite. We extend the notion of VC-dimension from natural numbers to ordinals as follows.

Definition 2. Let X be the set of nonempty subsets of $\mathcal{D}$ that $\mathcal{W}$ shatters. Let R be the restriction of $\supset$ to X. The *VC-dimension of* $\mathcal{P}$ is equal to the length of R if R is well-founded, and undefined otherwise.

As an intuitive interpretation, the VC-dimension of $\mathcal{P}$ is determined by the following game, where we assume for simplicity that $\mathcal{D}$ is infinite. Consider two players Anke and Boris. Anke has to output an increasing sequence of nonempty finite subsets of $\mathcal{D}$ and Boris a decreasing sequence of ordinals.

- In round 1, Anke outputs a nonempty finite subset of $\mathcal{D}$ and Boris outputs α.
- If in the n-th round Anke has output a finite set $D \subseteq \mathcal{D}$ and Boris has output a nonnull ordinal β, then the following is done in round $n+1$.
  - Anke outputs a finite set E with $D \subset E \subseteq \mathcal{D}$.
  - Boris outputs an ordinal strictly below β.

The game terminates after Boris has output 0. If Anke's last set is shattered, Anke wins the game. Otherwise Boris wins the game. The VC-dimension of $\mathcal{P}$ is the smallest nonnull ordinal α for which Boris has a winning-strategy. If this ordinal does not exist, the VC-dimension is undefined.

Given a structure $\mathfrak{M}$, we call *environment* (*in* $\mathcal{P}$) an infinite sequence e of members of $\mathcal{D} \cup \{\sharp\}$ such that for all $\varphi \in \mathcal{D}$, φ occurs in e iff $\varphi \in \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$. So environments correspond to texts when $\mathcal{D}$ is the set of closed atoms, and to informants when $\mathcal{D}$ is the set of closed literals. Identification in the limit and the corresponding notion of complexity are defined next.

Definition 3. An *identifier* (*for* $\mathcal{P}$) is a partial function from $(\mathcal{D} \cup \{\sharp\})^\star$ into $\{\mathrm{Diag}_{\mathcal{D}}(\mathfrak{M}) : \mathfrak{M} \in \mathcal{W}\}$. Let an identifier f be given.

- We say that a member σ of $(\mathcal{D} \cup \{\sharp\})^\star$ is *consistent in* $\mathcal{W}$ just in case there exists $\mathfrak{M} \in \mathcal{W}$ such that for all $\varphi \in \mathcal{D}$ that occur in σ, $\mathfrak{M} \models \varphi$.
- We say that f is *is successful* (*in* $\mathcal{P}$) iff for every $\mathfrak{M} \in \mathcal{W}$ and for every environment e for $\mathfrak{M}$, $f(e_{|k}) = \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$ for almost every $k \in \mathbb{N}$.
- Let X be the set of all $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ such that σ is consistent in $\mathcal{W}$ and $f(\tau) =\downarrow$ for some initial segment τ of σ. We denote by R_f the binary relation over X such that for all $\sigma, \tau \in X$, $R_f(\sigma, \tau)$ holds iff $\tau \subset \sigma$ and $f(\sigma) \neq f(\tau)$.

The *identification complexity of* $\mathcal{P}$ is the least ordinal of the form $|R_f|$, where f is an identifier that is successful in $\mathcal{P}$ and R_f is well-founded; if such an f does not exist, the identification complexity of $\mathcal{P}$ is undefined.

If the identification complexity of $\mathcal{P}$ is equal to nonnull ordinal β, then the $\mathcal{D}$-diagrams of the members of $\mathcal{W}$ are identifiable in the limit with less than β mind changes. (Note that the usual notion of mind change complexity considers the least ordinal β such that *at most*, rather than *less than*, β mind changes are sometimes necessary for the procedure to converge [1, 2, 7]. There are good theoretical reasons for preferring the 'less than' formulation.) Note that if some identifier is successful in $\mathcal{P}$, then there are countably many $\mathcal{D}$-diagrams of members of $\mathcal{W}$ only. The following is a characterization of identification complexity based on a generalization Angluin's finite telltale characterization of learnability in the limit [3].

Proposition 4. *The identification complexity of* $\mathcal{P}$ *is defined and equal to nonnull ordinal* β *iff there exists a sequence* $(\beta_{\mathfrak{M}})_{\mathfrak{M} \in \mathcal{W}}$ *of ordinals smaller than* β *and for all* $\mathfrak{M} \in \mathcal{W}$*, there exists a finite* $A_{\mathfrak{M}} \subseteq \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$ *such that for all* $\mathfrak{N} \in \mathcal{W}$*:*

$$(*) \quad \text{if } A_{\mathfrak{M}} \subseteq \mathrm{Diag}_{\mathcal{D}}(\mathfrak{N}) \text{ and } \beta_{\mathfrak{M}} \leq \beta_{\mathfrak{N}} \text{ then } \mathrm{Diag}_{\mathcal{D}}(\mathfrak{N}) = \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M}).$$

The next notion of complexity we define is based on selectors. Intuitively, a selector for $\mathcal{P}$ is a procedure that, for any given $\mathfrak{M} \in \mathcal{W}$, selects in the limit all formulas from $\mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$ or their negations, and is correct cofinitely many times. In order to be able to achieve this, the selector needs to receive for every selected formula φ, a feedback whether φ is or is not true in $\mathfrak{M}$. Thus one can represent the selector as a partial function where at stage n the input is from $\{0,1\}^n$ and represents the feed-back on the previous selections.

Definition 5. A *selector* (*for* $\mathcal{P}$) is a partial function f from $\{0,1\}^\star$ into $\mathcal{D}$. Let a selector f be given.

- We say that a member σ of $\{0,1\}^\star$ is *consistent with* $\mathcal{W}$ *and* f just in case there exists $\mathfrak{M} \in \mathcal{W}$ such that for all $\tau \in \{0,1\}^\star$ and $i \in \{0,1\}$ with $\tau i \subseteq \sigma$, $f(\tau) =\downarrow$, and $\mathfrak{M} \models f(\tau)$ iff $i = 1$.
- We say that f is *is successful* (*in* $\mathcal{P}$) iff for every $\mathfrak{M} \in \mathcal{W}$, there is a string t of finitely many 0's and infinitely many 1's such that every finite initial segment of t is consistent with $\mathcal{W}$ and f and for all $\varphi \in \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$, either $f(\sigma) = \varphi$ or $f(\sigma) = \neg\varphi$ for some finite initial segment σ of t.
- Let X be the set of all $\sigma \in \{0,1\}^\star$ that are consistent with $\mathcal{W}$ and f, and that end with a 0. We denote by R_f the binary relation over X such that for all $\sigma, \tau \in X$, $R_f(\sigma, \tau)$ holds iff $\tau \subset \sigma$.

The *selective complexity of* $\mathcal{P}$ is the least ordinal of the form $|R_f|$, where f is a selector that is successful in $\mathcal{P}$ and R_f is well-founded; if such a f does not exist, the selective complexity of $\mathcal{P}$ is undefined.

The last notion of complexity we define is based on predictors. Whereas selectors have control over the formulas the truth of which they want to predict, a predictor has to take the members of $\mathcal{D}$ as they come.

Definition 6. A *predictor* (*for* $\mathcal{P}$) is a partial function f from $(\mathcal{D} \times \{0,1\})^\star \times \mathcal{D}$ into $\{0,1\}$. Let a predictor f be given.

- Let $\sigma = ((\varphi_0, \epsilon_0), \ldots, (\varphi_p, \epsilon_p)) \in (\mathcal{D} \times \{0,1\})^\star$ be given. We say that σ is *consistent with* $\mathcal{W}$ *and* f iff there exists $\mathfrak{M} \in \mathcal{W}$ such that for all $i \leq p$:
  - $f(\sigma_{|i}, \varphi_i) =\downarrow$, and
  - $\mathfrak{M} \models \varphi_i$ iff $\epsilon_i = 1$

  If σ is consistent with $\mathcal{W}$ and f, we call the number of $i \leq p$ such that $f(\sigma_{|i}, \varphi_i) \neq \epsilon_i$, *the number of mispredictions that* f *makes on* σ (*in* $\mathcal{P}$).
- We say that f is *successful* (*in* $\mathcal{P}$) iff for every member $\mathfrak{M}$ of $\mathcal{W}$ and every $t \in (\mathcal{D} \times \{0,1\})^{\mathbb{N}}$, the following holds. Assume that every finite initial segment of t is consistent with $\mathcal{W}$ and f. Then there exists $n \in \mathbb{N}$ such that for all finite initial segments σ of t, f makes at most n mispredictions on σ.
- Let X be the set of all $\sigma \in (\mathcal{D} \times \{0,1\})^\star$ that are consistent with $\mathcal{W}$ and f, and on which f makes at least one misprediction. We denote by R_f the binary relation over X such that for all $\sigma, \tau \in X$, $R_f(\sigma, \tau)$ holds iff $\tau \subset \sigma$ and f makes more mispredictions on σ than on τ.

The *predictive complexity of* $\mathcal{P}$ is the least ordinal of the form $|R_f|$, where f is a predictor that is successful in $\mathcal{P}$ and R_f is well-founded; if such an f does not exist, the predictive complexity of $\mathcal{P}$ is undefined.

Clearly, if the predictive complexity of $\mathcal{P}$ is defined, then the selective complexity of $\mathcal{P}$ is defined, and at most equal to the former. The three notions of complexity we have introduced can, similarly to VC-dimension, be interpreted in terms of the outcome of a game between Anke and Boris. For instance, for predictive complexity, Anke selects formulas from $\mathcal{D}$ and Boris has to make a prediction whether the formula holds (in the unknown world $\mathfrak{M}$) or not. Anke tells Boris whether the prediction is correct. If not, Boris has to count down an ordinal counter. The predictive complexity is then the least ordinal to start with for which Boris has a winning strategy.

4 Relationships between the complexity measures

Let us first introduce an alternative definition of the selective complexity, followed by a definition and a lemma, that are just technical tools for proving some of the main propositions.

Notation 7. *Let $\alpha \in \mathrm{Ord}$ be given, and suppose that Γ_β has been defined for all $\beta < \alpha$. Let Γ_α be the set of all subsets U of $\mathcal{W}$ such that for all $\varphi \in \mathcal{D}$, $U \cap \mathrm{Mod}(\varphi)$ or $U \cap \mathrm{Mod}(\neg\varphi)$ belongs to $\bigcup_{\beta<\alpha} \Gamma_\beta \cup \{\emptyset\}$.*

Property 8. *The* predictive complexity of $\mathcal{P}$ *is defined iff $\mathcal{W} \in \bigcup_{\alpha\in\mathrm{Ord}} \Gamma_\alpha$; in this case it is equal to the least ordinal α with $\mathcal{W} \in \Gamma_\alpha$.*

Definition 9. A *shatterer (for $\mathcal{P}$)* is a sequence $(Y_\alpha)_{\alpha\in\mathrm{Ord}}$ of sets of subsets of $\mathcal{D}$ with the following properties.

- $\mathcal{W}$ shatters all members of $\bigcup_{\alpha\in\mathrm{Ord}} Y_\alpha$;
- for all $\alpha, \beta \in \mathrm{Ord}$ with $\beta \leq \alpha$ and for all $D, E \subseteq \mathcal{D}$ with $D \subseteq E$, if $E \in Y_\alpha$ then $D \in Y_\beta$;
- for all $\alpha, \beta \in \mathrm{Ord}$ with $\beta < \alpha$ and for all $E \in Y_\alpha$, Y_β contains a proper superset of E.

Lemma 10. *Let a sequence $(M_\alpha)_{\alpha\in\mathrm{Ord}}$ be inductively defined as follows. For all $\alpha \in \mathrm{Ord}$, M_α is the set of all finite subsets D of $\mathcal{D}$ such that $\mathcal{W}$ shatters D and for all $\beta < \alpha$, M_β contains a proper superset of D.*

- *$(M_\alpha)_{\alpha\in\mathrm{Ord}}$ is a shatterer.*
- *For all shatterers $(Y_\alpha)_{\alpha\in\mathrm{Ord}}$ and $\alpha \in \mathrm{Ord}$, $Y_\alpha \subseteq M_\alpha$.*
- *There exists a least ordinal α with $M_\alpha = M_{\alpha+1} \cup \{\emptyset\}$; the VC-dimension of $\mathcal{P}$ is defined and equal to α if $M_\alpha = \{\emptyset\}$, and undefined otherwise.*

We are now in a position to state and prove one of the main results of the paper, which relates VC-dimension to predictive complexity.

Proposition 11. *Suppose that $\mathcal{D}$ is closed under boolean operators. Then the VC-dimension of $\mathcal{P}$ is defined iff the predictive complexity of $\mathcal{P}$ is defined; moreover, if they are defined then they are equal.*

Proof. Given $\alpha \in \mathrm{Ord}$, let $\mathit{\Gamma}_\alpha$ denote the set of subsets of $\mathcal{W}$ as defined in Notation 7. For all $\alpha \in \mathrm{Ord}$, set $Z_\alpha = \{U \subseteq \mathcal{W} : U \notin \bigcup_{\beta<\alpha} \mathit{\Gamma}_\beta \cup \{\emptyset\}\}$. For all ordinals α, let Y_α be the set of all finite subsets E of $\mathcal{D}$ such that for all $D \subseteq E$, $\{\mathfrak{M} \in \mathcal{W} : \mathrm{Diag}_E(\mathfrak{M}) = D\} \in Z_\alpha$.

We show that $(Y_\alpha)_{\alpha\in\mathrm{Ord}}$ is a shatterer. Using the facts that $\emptyset \notin Z_\alpha$ and $\mathit{\Gamma}_\beta \subseteq \mathit{\Gamma}_\alpha$ for all $\alpha, \beta \in \mathrm{Ord}$ with $\beta \leq \alpha$, it is immediately verified that the first two conditions in Definition 9 are satisfied. For the third condition, let $\alpha, \beta \in \mathrm{Ord}$ with $\beta < \alpha$ and $E \in Y_\alpha$ be given. Given $D \subseteq E$, set $\varphi_D = \bigwedge D \wedge \neg \bigvee(E \setminus D)$. By the definition of Y_α, $\mathrm{Mod}_{\mathcal{W}}(\varphi_D) \in Z_\alpha$. Moreover, $\mathrm{Mod}_{\mathcal{W}}(\varphi_D) \notin \mathit{\Gamma}_\beta$. Hence there exists $\psi_D \in \mathcal{D}$ such that $\mathrm{Mod}_{\mathcal{W}}(\varphi_D \wedge \psi_D) \in Z_\beta$ and $\mathrm{Mod}_{\mathcal{W}}(\varphi_D \wedge \neg\psi_D) \in Z_\beta$. Set $\varphi = \bigvee\{\varphi_D \wedge \psi_D : D \subseteq E\}$. It is immediately verified that $E \cup \{\varphi\} \in Y_\beta$.

Assume that the predictive complexity of $\mathcal{P}$ is undefined. Let an ordinal α be given. Then $\mathcal{W} \notin \mathit{\Gamma}_\alpha$, hence $\mathcal{W} \in Z_{\alpha+1}$, hence $\emptyset \in Y_{\alpha+1}$. This implies that Y_α contains a non-empty set, and by Lemma 10, the VC-dimension of $\mathcal{P}$ is not equal to α. It follows that the VC-dimension of $\mathcal{P}$ is undefined.

Assume that the predictive complexity of $\mathcal{P}$ is defined and equal to ordinal γ. So $\mathcal{W} \in \mathit{\Gamma}_\gamma \cap Z_\gamma$. Let $\varphi \in \mathcal{D}$ be given, and suppose that $\{\varphi\} \in Y_\gamma$. Then both $\mathrm{Mod}_{\mathcal{W}}(\varphi)$ and $\mathrm{Mod}_{\mathcal{W}}(\neg\varphi)$ are members of Z_γ, which is in contradiction with $\mathcal{W} \in \mathit{\Gamma}_\gamma$. Hence $Y_\gamma = \{\emptyset\}$, and by Lemma 10 again, the VC-dimension of $\mathcal{P}$ is at least equal to γ. It is easily verified that the VC-dimension of $\mathcal{P}$ is defined and at most equal to γ, which completes the proof of the proposition.

The assumption that $\mathcal{D}$ is closed under negation is essential in Proposition 11, as shown in the next result.

Proposition 12. *For every nonnull $n \in \mathbb{N}$, there exists a finite vocabulary V and a finite set B of formulas over V with the following property. Suppose that $\mathcal{V} = V$, $\mathcal{W}$ is the set of Herbrand models of B, and $\mathcal{D}$ is the set of positive quantifier-free sentences. Then the VC-dimension of $\mathcal{P}$ is equal to 1 and the predictive complexity of $\mathcal{P}$ is equal to $\omega \times n$.*

Proof. Let a nonnull $n \in \mathbb{N}$ be given. Let $\mathcal{V}$ consist of a unary function symbol s, a binary predicate symbol $<$, an n-ary predicate symbol P and equality. Let B consist of the following formulas, which express that P is a predicate on $\mathbb{N}^n$ that is downward closed for the lexicographic ordering.

- $\forall x(x < s(x)) \wedge \forall xyz((x < y \wedge y < z) \rightarrow (x < z)) \wedge \forall xy(x < y \rightarrow \neg(y < x))$;
- $\forall x_1 \ldots x_n y_1 \ldots y_n((\bigvee_{i<n}(y_1 = x_1 \wedge \ldots \wedge y_i = x_i \wedge y_{i+1} < x_{i+1}) \wedge P(x_1, \ldots, x_n)) \rightarrow P(y_1, \ldots, y_n))$.

Clearly, the VC-dimension of $\mathcal{P}$ is equal to 1. It can be verified that the predictive complexity of $\mathcal{P}$ is equal to $\omega \times n$.

Denote by $\mathbb{W}$ the topological space over $\mathcal{W}$ generated by the sets of the form $\mathrm{Mod}_{\mathcal{W}}(\varphi)$, where φ varies over $\mathcal{D}$. To establish relationships between predictive, selective, and identification complexities, we usually need to assume that $\mathbb{W}$ is compact. This is the case for the next result.

Proposition 13. *Assume that $\mathcal{D}$ is closed under negation and that $\mathbb{W}$ is compact. If the identification complexity of $\mathcal{P}$ is defined and equal to ordinal α, then the predictive complexity of $\mathcal{P}$ is defined and at most equal to $\omega \times \alpha$.*

Proof. Suppose that the identification complexity of $\mathcal{P}$ is defined and equal to ordinal α. Choose a canonical identifier f for $\mathcal{P}$. Let a predictor h be defined as follows. Given a member $\sigma = ((\varphi_1, \epsilon_1), \ldots, (\varphi_n, \epsilon_n))$ of $(\mathcal{D} \times \{0,1\})^\star$, put $\widehat{\sigma} = (\psi_1, \ldots, \psi_n)$ where for all nonnull $i \leq n$, $\psi_i = \varphi_i$ if $\epsilon_i = 1$ and $\psi_i = \neg\varphi_i$ otherwise. Let $\sigma \in (\mathcal{D} \times \{0,1\})^\star$ and $\varphi \in \mathcal{D}$ be given. If $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\varphi)$ then $h(\sigma, \varphi) = 1$. If $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\neg\varphi)$ then $h(\sigma, \varphi) = 0$. If neither $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\varphi)$ nor $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\neg\varphi)$, and if $f(\widehat{\sigma})$ is defined and is a model of φ then $h(\sigma, \varphi) = 1$. Otherwise $h(\sigma, \varphi) = 0$. It is immediately verified that h is successful on $\mathcal{P}$. The proof of the proposition is completed if we show that the length of R_h is smaller than $\omega \times \alpha$. Let a member of the field of R_h of the form $\sigma \star (\varphi, 0)$ be given. Let X be the set of all $\tau \in (\mathcal{D} \cup \{\sharp\})^\star$ such that $\mathrm{rng}(\widehat{\sigma}) \subseteq \mathrm{rng}(\tau)$, $\{\varphi, \neg\varphi\} \cap \mathrm{rng}(\tau) \neq \emptyset$, τ is consistent in $\mathcal{W}$, and $f(\tau) =\downarrow$. By the choice of f, and since the restriction of $\mathbb{W}$ to $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma}))$ is compact, there exists a finite $F \subseteq X$ such that $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) = \bigcup\{\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\tau)) : \tau \in F\}$. Let n be the sum of the lengths of the members of F. It is easy to verify that $\rho_{R_h}(\sigma \star (\varphi, 0)) \leq \omega \times \sup_{\tau \in F} \rho_{R_f}(\tau) + n$. This implies immediately that R_h is at most equal to $\omega \times \alpha$, as wanted.

When $\mathcal{D}$ is closed under negation, identification and selective complexities are very close concepts, as shown next.

Proposition 14. *Suppose that $\mathcal{D}$ is closed under negation. If the selective complexity of $\mathcal{P}$ is defined and equal to ordinal α, then the identification complexity of $\mathcal{P}$ is defined and at most equal to $\alpha + 1$.*

Proof. Let a selector g be successful in $\mathcal{P}$ and such that the length of R_g is defined and equal to ordinal α. Without loss of generality, we can assume that for all $\sigma, \tau \in \{0,1\}^\star$, if $\sigma \subset \tau$ then $g(\sigma) \neq g(\tau)$. Let an identifier f be defined as follows. Let $\sigma = (\varphi_0, \ldots, \varphi_n) \in (\mathcal{D} \cup \{\sharp\})^\star$ be such that $\mathrm{cnt}(\sigma)$ is consistent in $\mathcal{W}$. We define a $p \in \mathbb{N}$, a member $\widehat{\sigma}$ of $\{0,1\}^\star$ and a finite sequence $(\psi_i)_{i<p}$ of members of $\mathcal{D}$ as follows. Let $i \in \mathbb{N}$ be given, and suppose that ψ_j has been defined for all $j < i$. Let $\tau = (\epsilon_0, \ldots, \epsilon_{i-1})$ be the unique member of $\{0,1\}^\star$ such that for all nonnull $j < i$, $\epsilon_j = 1$ if ψ_j occurs in σ, and $\epsilon_j = 0$ if $\neg\psi_j$ occurs in σ. If $g(\tau) =\downarrow$ and either $g(\tau)$ or $\neg g(\tau)$ occurs in σ, then $\psi_i = g(\tau)$. Otherwise $p = i$ and we set $\widehat{\sigma} = \tau$. Let $f(\sigma)$ consist of the set of all formulas occurring in σ and of all formulas of the form $g(\widehat{\sigma} \star \eta)$ where $\eta \in \{1\}^\star$ and $g(\widehat{\sigma} \star \eta) =\downarrow$. Set $f(\sigma) = X$. It is easily verified that f is successful in $\mathcal{P}$ and that the length of R_f is defined at most equal to $\alpha + 1$. This completes the proof of the proposition.

Assuming that $\mathcal{D}$ is closed under boolean operators, $\mathcal{W}$ is countable and $\mathbb{W}$ is compact, we can then put together the results that have been obtained previously and derive new connections, resulting in a complete picture of the relationships between all notions of complexity.

Proposition 15. *Suppose that $\mathcal{D}$ is closed under boolean operators, $\mathcal{W}$ is countable and $\mathbb{W}$ is compact. Then there exists ordinals α, β such that:*

- $\omega \times \alpha \leq \beta \leq \omega \times (\alpha + 1)$;
- *the VC-dimension of $\mathcal{P}$ is equal to β;*
- *the predictive complexity of $\mathcal{P}$ is equal to β;*
- *the identification complexity of $\mathcal{P}$ is equal to $\alpha + 1$;*
- *the selective complexity of $\mathcal{P}$ is equal to α or $\alpha + 1$.*

Proof. We first show that the VC-dimension of $\mathcal{P}$ is defined, which together with Propositions 11 and 14, implies immediately that the predictive complexity of $\mathcal{P}$, the identification complexity of $\mathcal{P}$, and the selective complexity of $\mathcal{P}$ are also defined. Suppose for a contradiction that there exists an infinite subset X of $\mathcal{D}$ such that for all finite $E \subseteq X$ and $D \subseteq E$, $D = \mathrm{Diag}_E(\mathfrak{M})$ for some $\mathfrak{M} \in \mathcal{W}$. Since $\mathbb{W}$ is compact, it follows that for all $Y \subseteq X$, $Y = \mathrm{Diag}_X(\mathfrak{M})$ for some $\mathfrak{M} \in \mathcal{W}$, which contradicts the assumption that $\mathcal{W}$ is countable. Thus every infinite subset of $\mathcal{D}$ has a finite subset that $\mathcal{W}$ does not shatter, and the VC-dimension of $\mathcal{P}$ is defined, as wanted.

We now show that the identification complexity of $\mathcal{P}$ is a successor ordinal. Choose a canonical identifier f for $\mathcal{P}$. It suffices to show that the length of R_f is not a limit ordinal. Suppose otherwise for a contradiction. Let X be the set of all finite subsets D of $\mathcal{D}$ such that $f(D) = \downarrow$. Since $\mathbb{W}$ is compact and f is successful in $\mathcal{P}$, there exists a finite subset F of X such that $\mathcal{W} = \bigcup\{\mathrm{Mod}_{\mathcal{W}}(D) : D \in F\}$. This implies that $|R_f|$ is equal to $\sup(\rho_{R_f}(\sigma) + 1 : D \in F)$, hence is smaller than $|R_f|$ since F is finite and $|R_f|$ is a limit ordinal. Contradiction.

Denote by $\alpha + 1$ the identification complexity of $\mathcal{P}$, and by β the predictive complexity of $\mathcal{P}$. By Proposition 11, the VC-dimension of $\mathcal{P}$ is equal to β. By Proposition 14, the selective complexity of $\mathcal{P}$ is at least equal to α. We show that the selective complexity of $\mathcal{P}$ is at most equal to $\alpha + 1$. Choose a set-driven identifier f such that f is successful in $\mathcal{P}$ and the length of R_f is defined and equal to ordinal $\alpha + 1$. Fix an enumeration $(\varphi_i)_{i \in \mathbb{N}}$ of $\mathcal{D}$. We define by induction a sequence $(h_i)_{i \in \mathbb{N}}$ of selectors with finite domains such that for all $i, j \in \mathbb{N}$, h_i and h_j have disjoint domains. Let $i \in \mathbb{N}$ be given, and suppose that h_j has been defined for all $j < i$. If $i = 0$ put $Z = \{()\}$. If $i \neq 0$, let Z be the set of all $\subseteq$-maximal members of the domain of h_{i-1}. Let $\sigma \in Z$ be given. If $\sigma = ()$ let $\widehat{\sigma} = \sigma$. If σ is nonempty and equal to $(\epsilon_0, \ldots, \epsilon_k)$, denote by $\widehat{\sigma}$ the sequence $(\psi_0, \ldots, \psi_k)$ of members of $\mathcal{D}$ such that for all $j \leq k$, $\psi_j = f(\sigma_{|j})$ if $\epsilon_j = 1$, and $\psi_j = \neg f(\sigma_{|j})$ otherwise. If $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\varphi_i)$ then set $h_i(\sigma) = \varphi_i$. If $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\neg\varphi_i)$ then set $h_i(\sigma) = \neg\varphi_i$. Suppose that neither $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\varphi_i)$ nor $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) \subseteq \mathrm{Mod}(\neg\varphi_i)$. Let X_σ be the set of all $\tau \in (\mathcal{D} \cup \{\sharp\})^\star$ such that $\mathrm{rng}(\widehat{\sigma}) \subseteq \mathrm{rng}(\tau)$, $\{\varphi_i, \neg\varphi_i\} \cap \mathrm{rng}(\tau) \neq \emptyset$, τ is consistent in $\mathcal{W}$, and $f(\tau) = \downarrow$. By the choice of f, and since the restriction of $\mathbb{W}$ to $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma}))$ is compact, there exists a finite $F_\sigma \subseteq X_\sigma$ such that $\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\widehat{\sigma})) = \bigcup\{\mathrm{Mod}_{\mathcal{W}}(\mathrm{rng}(\tau)) : \tau \in F_\sigma\}$. Without loss of generality, we can assume that there exists $\psi_0, \ldots, \psi_k \in \mathcal{D}$ such that for all for all $\tau \in F_\sigma$ and for all $j \leq k$, either ψ_j or $\neg\psi_j$ occurs in τ, and every formula that occurs in τ is of the form ψ_j or $\neg\psi_j$ for some $j \leq k$. Let n be the cardinality of F_σ, and fix an enumeration $(D_p)_{p<n}$ of F_σ. Note that for all distinct $p, p' < n$, $\mathrm{Mod}_{\mathcal{W}}(D_p) \cap \mathrm{Mod}_{\mathcal{W}}(D_{p'}) = \emptyset$. Given $m < n$, set $\overline{\sigma}^m = 1^{m-1}01^{n-m-1}$. Now for all $\sigma \in Z$, $m < n$ and $p < n$, set $h_i(\sigma \star \overline{\sigma}^m_{|p}) = \neg \bigwedge D_p$; for all $\sigma \in Z$ and $m < n$, set $h_i(\sigma \star \overline{\sigma}^m) = \varphi_i$ if $\bigwedge D_m \models \varphi_i$, and $h_i(\sigma \star \overline{\sigma}^m) = \neg\varphi_i$ otherwise. Set $h = \bigcup_{i\in\mathbb{N}} h_i$. It is easily verified that h succeeds in $\mathcal{P}$, and that the length of R_h is defined and at most equal to $\alpha + 1$.

By Proposition 13, $\beta \leq \omega \times (\alpha + 1)$. So to complete the proof of the proposition, it suffices to show that $\omega \times \alpha \leq \beta$. Remember the definition of the sequence $(\mathit{\Gamma}_\alpha)_{\alpha\in\mathrm{Ord}}$ from Notation 7. By Property 8, $\mathcal{W} \in \mathit{\Gamma}_\beta$. Define an identifier f as follows. Let $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ be given. Let γ be the least ordinal with $\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma)) \in \mathit{\Gamma}_\gamma$.

- If there exists $\tau \subset \sigma$ such that $f(\tau)$ is defined and equal to the $\mathcal{D}$-diagram of a model of $\mathrm{cnt}(\sigma)$ in $\mathcal{W}$, then $f(\sigma) = f(\tau)$.
- Otherwise, and if there exists $\mathfrak{M} \in \mathcal{W}$ such that for all $\varphi \in \mathcal{D}$, $\mathfrak{M}$ is a model of φ iff $\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma) \cup \{\varphi\}) \notin \bigcup_{\gamma'<\gamma} \mathit{\Gamma}_{\gamma'}$, then $f(\sigma) = \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$.
- Otherwise $f(\sigma) =\uparrow$.

Let $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ be such that $f(\sigma) =\downarrow$, and let γ be the least ordinal with $\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma)) \in \mathit{\Gamma}_\gamma$. Suppose for a contradiction that γ is neither 0 nor a limit ordinal. It then follows from the definition of $\mathit{\Gamma}_\gamma$ that there exists $\varphi \in \mathcal{D}$ such that $\mathit{\Gamma}_{\gamma-1}$ contains neither $\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma) \cup \{\varphi\})$ nor $\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma) \cup \{\neg\varphi\})$, which is impossible by the definition of f. Let λ be the number of limit ordinals less than or equal to β. We infer easily from the previous remark that $\alpha+1 \leq \lambda+1$, hence $\omega \times \alpha \leq \omega \times \lambda \leq \beta$, and we are done.

We illustrate the previous results, especially the bounds that have been obtained, with a few examples.

Example 16. Let $\mathcal{V}$ consists of $\overline{0}$, s, and a unary predicate symbol P. Let $\mathcal{W}$ be the set of Herbrand models of $\forall x(P(x) \wedge P(s(x)) \rightarrow P(s(s(x)))) \wedge \exists y(P(y) \wedge P(s(y)))$. Then the identification complexity of $\mathcal{P}$ is 1 but the predictive complexity of $\mathcal{P}$ is undefined.

Example 17. Assume that $\mathcal{D}$ is closed under Boolean operators. If the VC-dimension of $\mathcal{P}$ is a nonnull $n \in \mathbb{N}$ then both the identification and the selective complexity of $\mathcal{P}$ are 1, and the predictive complexity of $\mathcal{P}$ is n.

Proposition 18. *Suppose that $\mathcal{V}$ consists of $\overline{0}$, s, and a unary predicate symbol P. Set $\mathcal{D} = \{P(\overline{n}) : n \in \mathbb{N}\}$. Assume that $\mathcal{W}$ is the set of Herbrand structures whose $\mathcal{D}$-diagrams are $\{\overline{nk} : n \in \mathbb{N}\}$ for $k \in \mathbb{N}$. Then the identification complexity of $\mathcal{P}$ is 2, and both the VC-dimension and the predictive complexity of $\mathcal{P}$ are ω.*

Proof. It is easily verified that an identifier can be successful in $\mathcal{P}$ by first hypothesizing $\{P(\overline{0})\}$, and in the case of an environment for the structure whose $\mathcal{D}$-diagram is $T = \{P(\overline{nk}) : n \in \mathbb{N}\}$ for some nonnull $k \in \mathbb{N}$, by eventually changing $\{P(\overline{0})\}$ to T. If follows that the identification complexity of $\mathcal{P}$ is 2. It is easily verified that the VC-dimension and the predictive complexity of $\mathcal{P}$ are at most equal to ω. To see that they are at least equal to ω, let a nonnull $n \in \mathbb{N}$ be given, and let $p_0, p_1, \ldots, p_{2^n-1}$ be an enumeration of the first 2^n prime numbers. For all $k < n$, let q_k be the product of all p_i, $0 \leq i < 2^n$, such that the $(k+1)st$ bit in the binary representation of i is equal to 1. It is immediately verified that $\mathcal{W}$ shatters $\{P(\overline{q_0}), P(\overline{q_1}), \ldots, P(\overline{q_{n-1}})\}$, implying that the VC-dimension of $\mathcal{P}$ is at least equal to n, hence the predictive complexity of $\mathcal{P}$ is also at least equal to n.

Example 19. Suppose that $\mathcal{V}$ consists of $\overline{0}$, s, and a unary predicate symbol P. Assume that $\mathcal{W}$ is the set of Herbrand structures $\mathfrak{M}$ such that for all $n \in \mathbb{N}$, if $\mathfrak{M} \models P(\overline{n})$ then there exists at most n natural numbers m such that $\mathfrak{M} \models P(\overline{m})$. Suppose that $\mathcal{D}$ is the set of quantifier free sentences. Then the identification complexity of $\mathcal{P}$ is $\omega+1$, the selective complexity of $\mathcal{P}$ is ω, and the VC-dimension of $\mathcal{P}$ is ω^2.

Note that if in the previous example, we remove from $\mathcal{W}$ the structure $\mathfrak{M}$ such that $\mathfrak{M} \models \forall x \neg P(x)$—resulting in a noncompact topological space $\mathbb{W}$—, then the identification complexity becomes ω; but in both cases, the traditional notion of mind change complexity is ω.

Proposition 20. *Let a countable ordinal α be given. There exists a finite vocabulary V and two sets K and E of sentences over V such that if $\mathcal{V} = V$, $\mathcal{D} = E$, and $\mathcal{W}$ is the set of Herbrand models of K, then the identification complexity of $\mathcal{P}$ is $\alpha + 1$, the selective complexity of $\mathcal{P}$ is α, and both the VC-dimension and the predictive complexity of $\mathcal{P}$ are $\omega \times \alpha$.*

Proof. Suppose that $\mathcal{V}$ consists of a constant, a unary function symbol, a unary predicate symbol P, and a binary predicate symbol $\leq$. Identify the set X of closed terms with the set of ordinals smaller than α, writing $X = \{c_\beta : \beta < \alpha\}$. Let $\mathcal{D}$ be the closure of $\{P(c_\beta) : \beta < \alpha\}$ under boolean operators. Suppose that $\mathcal{W}$ is the set of Herbrand models of

$$\{(c_\beta, c_\gamma) : \beta, \gamma < \alpha,\ \beta \leq \gamma\} \cup \{\forall x \forall y((P(x) \land x \leq y) \to P(y))\}.$$

We define by induction, for all members σ of $(\mathcal{D} \cup \{\sharp\})^\star$, three ordinals: $\mathrm{counter}_\sigma$, above_σ, and below_σ. Set $\mathrm{counter}_{()} = \alpha$, $\mathrm{above}_{()} = \omega^\alpha$ and $\mathrm{below}_{()} = 0$. Let $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ and $x \in \mathcal{D} \cup \{\sharp\}$ be given, and suppose that $\mathrm{counter}_\sigma$, above_σ and below_σ have been defined. Let $\mathrm{above}_{\sigma\star x}$ be the minimal ordinal β such that c_β occurs in $\sigma \star x$ if such a β exists, and let $\mathrm{above}_{\sigma\star x}$ be ω^α otherwise. Let $\mathrm{below}_{\sigma\star x}$ be the maximal ordinal β such that c_β occurs in $\sigma \star x$ if such a β exists, and let $\mathrm{below}_{\sigma\star x}$ be 0 otherwise. If there exist (unique) ordinals β, γ such that $\beta < \mathrm{counter}_\sigma$, $\mathrm{above}_\sigma = \omega^\beta \times \gamma$ and $\mathrm{below}_\sigma + \omega^\beta \geq \mathrm{above}_\sigma$, then set $\mathrm{counter}_{\sigma\star x} = \beta$; otherwise set $\mathrm{counter}_{\sigma\star x} = \mathrm{counter}_\sigma$. Let f be the identifier defined as follows. Let $\sigma \in (\mathcal{D} \cup \{\sharp\})^\star$ be given. If $\mathrm{counter}_\sigma = \omega^\alpha$ then $f(\sigma)$ is the $\mathcal{D}$-diagram of the member $\mathfrak{M}$ of $\mathcal{W}$ such that $\mathfrak{M} \models \forall x \neg P(x)$; otherwise, $f(\sigma)$ is the $\mathcal{D}$-diagram of the member $\mathfrak{M}$ of $\mathcal{W}$ such that $\mathfrak{M} \models \forall x(P(x) \leftrightarrow x \geq c_{\mathrm{above}_\sigma})$. It is immediately verified that f is successful in $\mathcal{P}$, and that the length of R_f is defined and equal to $\alpha + 1$. It is also easily verified that the identification complexity of $\mathcal{P}$ cannot be smaller than $\alpha + 1$, and that the selective complexity of $\mathcal{P}$ is α.

Since $\mathbb{W}$ is a compact topological space, it follows from Propositions 11 and 15 that the VC-dimension and the predictive complexity of $\mathcal{P}$ are equal, and at most equal to $\omega \times \alpha$. So to complete the proof of the proposition, it suffices to show that the VC-dimension of $\mathcal{P}$ is at least equal to $\omega \times \alpha$. Suppose for a contraction that the VC-dimension of $\mathcal{P}$ is equal to ordinal α' with $\alpha' < \alpha$. The argument uses the presentation of the VC-dimension as a game between Anke and Boris. At round n, Anke defines a set of ordinals O_n and a formula ψ_n, then outputs a set E_n, before Boris outputs an ordinal β_n. Set $O_0 = \{0\}$, $\psi_0 = P(c_{\alpha'})$, $E_0 = \{\psi_0\}$, and $\beta_0 = \alpha'$. Let $n \in \mathbb{N}$ be given and suppose that O_n, ψ_n, E_n and β_n have been defined. Let ordinal γ and $m \in \mathbb{N}$ be such that $\beta_n = \omega \times \gamma + m$. Set $O_{n+1} = O_n \cup \{\delta + \omega^\gamma + 2^m : \delta \in O_n\}$. Let ψ_{n+1} be a formula expressing that the cardinality of the members x of O_{n+1} that have property P is odd. Let $E_{n+1} = \{\psi_0 \ldots, \psi_{n+1}\}$. Note that O_{n+1} expands O_n with exactly one ordinal between two ordinals in O_n with no ordinal in O_n in between, plus an ordinal greater than all ordinals in O_n. It is easily verified that for all $n \in \mathbb{N}$, $\mathcal{W}$ shatters E_n. Hence Anke wins the game, contradiction.

If the previous example were modified by taking $\{P(c_\beta) : \beta < \alpha\}$ as set of possible data, resulting in a new paradigm $\mathcal{P}'$, then the identification complexity of $\mathcal{P}'$ would clearly be equal to $\omega^\alpha + 1$. More generally, there exists a relationship between the identification complexities of two paradigms that differ in their sets of possible data—one set of data being closed under negation in one paradigm, and not in the other. This relationship is considered in the next proposition, and the previous example shows that the "exponential" difference between the identification complexities of both paradigms is almost as large as it can be. First note that Angluin's finite telltale condition takes the following form in the logical framework.

Lemma 21. *Some identifier is successful in $\mathcal{P}$ iff for all $\mathfrak{M} \in \mathcal{W}$, there exists a finite $A \subseteq \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$ such that no $\mathfrak{N} \in \mathcal{W}$ satisfies $A \subseteq \mathrm{Diag}_{\mathcal{D}}(\mathfrak{N}) \subset \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M})$.*

Proposition 22. *Suppose that $\mathcal{D}$ is the closure under negation of some set of sentences $\mathcal{D}'$, and set $\mathcal{P}' = (\mathcal{V}, \mathcal{D}', \mathcal{W})$. Assume that $\mathbb{W}$ is compact, the identification complexity of $\mathcal{P}$ is defined and equal to nonnull ordinal α, and some identifier is successful in $\mathcal{P}'$. Then the identification complexity of $\mathcal{P}'$ is defined and smaller than ω^α.*

Proof. The proof is trivial if $\{\mathrm{Diag}_{\mathcal{D}}(\mathfrak{M}) : \mathfrak{M} \in \mathcal{W}\}$ is finite, so suppose otherwise. Let $(\psi_i)_{i\in\mathbb{N}}$ be a repetition free enumeration of $\mathcal{D}'$. Let X be the set of finite sequences of the form $(\varphi_0, \ldots, \varphi_{n-1})$, $n \in \mathbb{N}$, where for all $i < n$, $\varphi_i = \psi_i$ or $\varphi_i = \neg\psi_i$, such that $\{\varphi_0, \ldots, \varphi_{n-1}\}$ is consistent in $\mathcal{W}$. Let f be a canonical identifier for $\mathcal{P}$. Let a finite subset E of $\mathcal{D}'$ be consistent in $\mathcal{W}$. Denote by U_E the set of all $\subseteq$-minimal members σ of X such that $f(\sigma)$ is defined and contains E. Suppose for a contradiction that U_E is infinite. By König's lemma, there exists a sequence $(\xi)_{i\in\mathbb{N}}$ of formulas such that for all $n \in \mathbb{N}$, $(\xi_0, \ldots, \xi_{n-1}) = \sigma_{|n}$ for some $\sigma \in U_E$. Note that $E \subseteq \mathrm{rng}(\sigma)$ for cofinitely many members σ of U_E, hence $E \subseteq \{\xi_i : i \in \mathbb{N}\}$. Moreover, for all distinct $\sigma, \tau \in U_E$, neither $\sigma \subseteq \tau$ nor $\tau \subseteq \sigma$, hence $\sigma \not\subset \{\xi_i : i \in \mathbb{N}\}$ for cofinitely many members σ of U_E. Hence $f((\xi_0, \ldots, \xi_n))$ is either undefined or does not contain E for infinitely many $n \in \mathbb{N}$. By compactness of $\mathbb{W}$, $\{\xi_i\}_{i\in\mathbb{N}}$ is the $\mathcal{D}$-diagram of a member of $\mathcal{W}$. But this is in contradiction with the fact that f is successful in $\mathcal{P}$. So we have verified that U_E is finite. Note that since f is successful in $\mathcal{P}$, $\mathrm{Mod}_{\mathcal{W}}(E)$ is included in $\bigcup\{\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\sigma)) : \sigma \in U_E\}$. Let $(\sigma_1, \ldots, \sigma_p)$ be a repetition free enumeration of U_E with $\rho_{R_f}(\sigma_1) \geq \ldots \geq \rho_{R_f}(\sigma_p)$. Set $\beta_E = \omega^{\rho_{R_f}(\sigma_1)} + \ldots + \omega^{\rho_{R_f}(\sigma_p)}$. Let finite $E, F \subseteq \mathcal{D}$ be such that $E \subseteq F$, $U_F \setminus U_E \neq \emptyset$, and U_F is nonempty. Let $\tau \in U_F \setminus U_E$ be given. Then there exists $i \in \{1, \ldots, p\}$ such that $\sigma_i \subset \tau$, $f(\sigma_i) \neq f(\tau)$, and $\sigma_i \notin U_F$. Let D be the set of all members of U_F that strictly extend σ_i. Clearly, $\rho_{R_f}(\sigma_i) > \rho_{R_f}(\gamma)$ for all $\gamma \in D$. Moreover, the factor $\omega^{\rho_{R_f}(\sigma_i)}$ in the sum β_E is replaced in the sum β_F by the factors $\omega^{\rho_{R_f}(\gamma)}$, with γ varying over D. Due to the powers of ω, it is easily verified that the extra factors $\omega^{\rho_{R_f}(\gamma)}$, $\gamma \in D$, 'weight less' than $\omega^{\rho_{R_f}(\sigma_i)}$. Since every $\sigma_i \in U_E$ is either also in U_F or replaced in E_F by a D with the properties just described, we conclude that $\beta_E > \beta_F$.

Fix a repetition free enumeration $(T_i)_{i\in\mathbb{N}}$ of $\{\mathrm{Diag}_{\mathcal{D}}(\mathfrak{M}) : \mathfrak{M} \in \mathcal{W}\}$. By Lemma 21, choose for all $i \in \mathbb{N}$ a finite subset A_i of T_i such that no member $\mathfrak{M}$ of $\mathcal{W}$ satisfies $A_i \subseteq \mathrm{Diag}_{\mathcal{D}}(\mathfrak{M}) \subset T_i$. Let Y be the set of all $\sigma \in (\mathcal{D}' \cup \{\sharp\})^\star$ such that $A_i \subseteq \mathrm{cnt}(\tau) \subseteq T_i$ for some $i \in \mathbb{N}$ and some initial segment τ of σ. For all $\sigma \in Y$, denote by $\widehat{\sigma}$ the $\subseteq$-maximal initial segment of σ such that $A_i \subseteq \mathrm{cnt}(\widehat{\sigma}) \subseteq T_i$ for some $i \in \mathbb{N}$. Now define an identifier f' for $\mathcal{P}'$ as follows. Let $\sigma \in (\mathcal{D}' \cup \{\sharp\})^\star$ be given. If $\sigma \notin Y$ then $f'(\sigma) =\uparrow$; otherwise, $f'(\sigma) = T_i$ where $i \in \mathbb{N}$ is least with $A_i \subseteq \mathrm{cnt}(\widehat{\sigma}) \subseteq T_i$. Let $\sigma, \tau \in (\mathcal{D}' \cup \{\sharp\})^\star$ be such that $\sigma \subset \tau$, and $f'(\sigma)$ and $f'(\tau)$ are defined but distinct. Then both σ and τ belong to Y. Clearly, there exists $i \in \mathbb{N}$ such that $\mathrm{cnt}(\sigma) \subseteq T_i$ but $\mathrm{cnt}(\tau) \not\subseteq T_i$, which implies that

$$T_i \in \bigcup\{\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\gamma)) : \gamma \in U_{\mathrm{cnt}(\tau)}\} \setminus \bigcup\{\mathrm{Mod}_{\mathcal{W}}(\mathrm{cnt}(\gamma)) : \gamma \in U_{\mathrm{cnt}(\sigma)}\}.$$

We infer that $U_{\mathrm{cnt}(\tau)}$ is nonempty and distinct from $U_{\mathrm{cnt}(\sigma)}$, which we know implies that $\beta_{\mathrm{cnt}(\widehat{\sigma})} > \beta_{\mathrm{cnt}(\widehat{\tau})}$. It then follows that the height of $R_{f'}$ is defined and at most equal to ω^α. Moreover, since $\mathbb{W}$ is compact, the reasoning in Proposition 15 shows that the identification complexity of $\mathcal{P}$ is not a limit ordinal, hence is smaller than ω^α, as wanted.

It was shown in [11] that every class of languages that is finitely identifiable from informants is also identifiable in the limit from texts. But such a relationship does not generalize to identifiability from informants with one mind change at most. Indeed, the class $\mathcal{C}$ consisting of $\mathbb{N}$ and all finite initial segments of $\mathbb{N}$ is not learnable in the limit from texts, as proved in [8], whereas $\mathcal{C}$ is clearly learnable from informants with one mind change at most. Cast into the logical framework, this provides an example of $\mathcal{V}$, $\mathcal{W}$, $\mathcal{D}$, and $\mathcal{D}'$ such that $\mathcal{D}$ is the closure of $\mathcal{D}'$ under negation, $\mathbb{W}$ is compact, the identification complexity of $\mathcal{P}$ is equal to 1, but no identifier is successful in $\mathcal{P}' = (\mathcal{V}, \mathcal{D}', \mathcal{W})$. This shows that Proposition 22 does not hold if the assumption that some identifier is successful in $\mathcal{P}$ is dropped.

5 Conclusion

.

In ideals paradigms of Inductive inference, finite tell tails conditions offer characterizations of identification in the limit or of classification, with or without a mind change bound. Assuming that the set of data is closed under boolean operators, the VC-dimension offers a characterization of prediction. An extra topological assumption of compactness enables to provide a complete picture of the relationship between VC-dimension and other notions of complexity, including mind change bound complexity.

References

1. Ambainis, A., Jain, S., Sharma, A.: *Ordinal mind change complexity of language identification.* Theoretical Computer Science. **220(2)** pp. 323-343 (1999)
2. Ambainis, A., Freivalds, R., Smith, C.: *Inductive Inference with Procrastination: Back to Definitions.* Fundamenta Informaticae. **40** pp. 1-16 (1999)
3. Angluin, D. *Inductive Inference of Formal Languages from Positive Data.* Information and Control **45** p. 117-135 (1980)
4. Ben-David, S., Gurvits, L.: *A note on VC-Dimension and Measure of Sets of Reals.* Combinatorics Probability and Computing 9, 391-405 (2000)
5. Ben-David, S., Jacovi, M.: *On Learning in the Limit and Non-Uniform* (ϵ, δ)*-Learning.* In Proceedings of the Sixth Conference on Computational Learning Theory. ACM Press pp. 209-217 (1993)
6. Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.: *Learnability and the Vapnik-Chervonenkis Dimension.* J. ACM **36(4)** pp. 929-965 (1989)
7. Freivalds, R., Smith, C.: *On the role of procrastination for machine learning.* Inform. Comput. **107(2)** pp. 237-271 (1993)
8. Gold, E.: *Language Identification in the Limit.* Information and Control. **10** (1967)
9. Martin, E., Sharma, A., Stephan, F.: *A General Theory of Deduction, Induction, and Learning.* In Jantke, K., Shinohara, A.: Proceedings of the Fourth International Conference on Discovery Science. Springer-Verlag, LNAI 2226 pp. 228-242 (2001)
10. Martin, E., Sharma, A., Stephan, F.: *Logic, Learning, and Topology in a Common Framework.* In Cesa-Bianchi, N., Numao, M., Reischuk, R.: Proc. of the 13th Intern. Conf. on Alg. Learning Theory. Springer-Verlag, LNAI 2533 pp. 248-262 (2002)
11. Sharma, A.: *A note on batch and incremental learnability.* Journal of Computer and System Sciences **56** pp. 272-276 (1998)
12. Valiant, L.: *A Theory of the Learnable.* Commun. ACM **27(11)** pp. 1134-1142 (1984)
13. Vapnik, V., Chervonenkis, A.: *On the uniform convergence of relative frequencies of events to their probabilities.* Theory of Probabilities and its Applications **16(2)** pp. 264-280 (1971)