Relations between Gold-style learning and query learning

Steffen Lange [a] and Sandra Zilles [b,*]

[a] *Fachbereich Informatik, Fachhochschule Darmstadt, Haardtring 100, 64295 Darmstadt, Germany*

[b] *DFKI GmbH, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany*

Abstract

Different formal learning models address different aspects of human learning. Below we compare *Gold-style learning*—modelling learning as a *limiting process* in which the learner may change its mind arbitrarily often before converging to a correct hypothesis—to *learning via queries*—modelling learning as a *one-shot process* in which the learner is required to identify the target concept with just one hypothesis. In the Gold-style model considered below, the information presented to the learner consists of positive examples for the target concept, whereas in query learning, the learner may pose a certain kind of queries about the target concept, which will be answered correctly by an oracle (called teacher).

Although these two approaches seem rather unrelated at first glance, we provide characterisations of different models of Gold-style learning (learning in the limit, conservative inference, and behaviourally correct learning) in terms of query learning. Thus we describe the circumstances which are necessary to replace limit learners by equally powerful one-shot learners. Our results are valid in the general context of learning indexable classes of recursive languages.

This analysis leads to an important observation, namely that there is a natural query learning type hierarchically in-between Gold-style learning in the limit and behaviourally correct learning. Astonishingly, this query learning type can then again be characterised in terms of Gold-style inference.

Key words: Inductive inference, query learning, formal languages, recursion theory

* corresponding author

Email addresses: s.lange@fbi.fh-darmstadt.de (Steffen Lange), sandra.zilles@dfki.de (Sandra Zilles).

1 Introduction

Undeniably, there is no formal scheme spanning all aspects of human learning. Thus each learning model analysed within the scope of learning theory addresses only special facets of our understanding of learning.

For example, Gold's [9] model of *identification in the limit* is concerned with learning as a limiting process of creating, modifying, and improving hypotheses about a target concept. These hypotheses are based upon positive examples [1] for the target concept sequentially offered as information. In the limit, given a gradually growing sequence of positive examples, the learner is supposed to stabilize on a correct guess, but during the learning process in general one will never know whether or not the current hypothesis is already correct. The reason is that at any time in the learning process, the learner has only seen a finite sequence of positive examples and may change its hypothesis upon the next example in the sequence to be presented. Here the ability to change its mind is a crucial feature of the learner.

In contrast to that, Angluin's [2,3] model of *learning with queries* focusses learning as a finite process of interaction between a learner and a teacher. The learner asks questions of a specified type about the target concept and the teacher—having the target concept in mind—answers these questions truthfully. After finitely many steps of interaction the learner is supposed to return its sole hypothesis—correctly describing the target concept. Here the crucial features of the learner are its ability to demand special information on the target concept and its restrictiveness in terms of mind changes. Since a query learner is required to identify the target concept with just a single hypothesis, we refer to this phenomenon as *one-shot learning*.

Note that, in contrast to Gold's model, where the input of the learner consists only of positive examples for the target concept, in query learning the information the learner gets may be of a different quality, so concerning the information resources, the preconditions of query learners differ from those of Gold-style learners. Another difference obviously lies in the constraints on the convergence of the learning process: while a Gold-style learner is only required to converge to a correct guess in the limit, query learners have to deliberately stop the learning process with a single and correct guess.

Our analysis concerns common features and relations between these two seemingly unrelated approaches, thereby focussing our attention on the identification of formal languages, ranging over indexable classes of recursive languages, as target concepts, see Angluin [1], Lange and Zeugmann [13], and Zeugmann

[1] Gold [9] also initiated a model of learning from both positive and negative examples, but we neglect this approach in the sequel.

and Lange [21]. In this context, our main focus will be on characterisations of Gold-style language learning in terms of learning via queries. Characterising different types of Gold-style language learning in such a way, we will point out interesting correspondences between the two models. Our results illustrate that a difference in the quality of the information resources can be traded for a difference in the requirements of convergence concerning the hypotheses. In particular, it is demonstrated how learners identifying languages in the limit can be replaced by one-shot query learners without any loss of learning power. That means, under certain circumstances the capability of limit learners is equal to that of one-shot learners using queries, or, in other words, each class of target languages learnable in Gold's model is learnable in the query model and vice versa.

Note that there is a trivial type of queries that could turn Gold-style limit-learners into one-shot learners: if a Gold-style learner M was additionally told by an oracle whether or not its own sequence of hypotheses has already passed the point of convergence for the target language, then it would be possible to learn each language identified in the limit by M with only a single hypothesis. But in such a model, the oracle would have to use knowledge about the specific learner M, which makes learning trivial. In the original query model, the oracle is independent from the learner and thus it is much more interesting, whether the capabilities of Gold-style learners can be featured by such a query learner. A first idea might be to use so-called equivalence queries: if the learner asks an equivalence query concerning some language L, then the teacher replies 'yes', if L equals the target language and 'no', otherwise. It is well-known and trivial, that each indexable class of recursive languages can be learned with equivalence queries with just a single hypothesis. Thus, in particular, each class learnable in the limit in Gold's model is also identifiable with a one-shot query learner using equivalence queries. But this result is again not very satisfying, since the capabilities of equivalence query learners outperform those of Gold-style learners in a trivial way, namely by being able to learn each conceivable target class. What is much more challenging in this context is the relation between Gold-style models and query models which are non-trivial concerning the learnability of indexable classes of languages.

Studying such relations between two different approaches to language learning may allow for transferring theoretically approved insights from one model to the other. In particular, our characterisations may serve as 'interfaces' between an analysis of query learning and an analysis of Gold-style learning through which proofs on either model can be simplified using features of the other.

The crucial question in this context is what abilities of the teacher are required to achieve the learning capabilities of Gold-style learners for query learners. In particular, it is of importance which types of queries the teacher is able to answer (and thus the learner is allowed to ask). This addresses two facets:

first, the kind of information inquired by the queries (we consider membership, restricted superset, and restricted disjointness queries) and second, the hypothesis space used by the learner to formulate its queries and hypotheses (we consider uniformly recursive, uniformly r. e., and uniformly K-r. e. families). Note that both aspects affect the demands on the teacher. In particular, most of the cases we consider in general require a very powerful teacher capable of answering undecidable questions.

Our characterisations reveal the corresponding necessary requirements that have to be made on the teacher. Thereby we formulate relations of the learning capabilities assigned to Gold-style learners and query learners in a quite general context, considering three variants of Gold-style language learning.

Our analysis will lead to the following observation: there is a natural inference type (learning via restricted superset queries in Gödel numberings) which lies *in-between* Gold-style learning in the limit from text and behaviourally correct Gold-style learning from text in Gödel numberings. That means that the capabilities of the corresponding learners lie in-between. Up to now, no such inference type has been known. Concerning the notions of inference, see Gold [9], Angluin [1], Zeugmann and Lange [21], and the preliminaries below.

This observation immediately raises the question, whether there is an analogue of this query learning type in terms of Gold-style learning and thus whether there is also a *Gold-style* inference type between learning in the limit and behaviourally correct learning. Indeed such a relation can be observed with conservative inference in Gödel numberings by learners using an oracle for the halting problem; see Stephan [20] for further results on learning with oracles. This also corroborates the impression that the relations between Gold-style learning and query learning studied here are not coincidental but quite fundamental.

As a byproduct of the proofs, we provide special indexable classes of recursive languages which can be learned in a behaviourally correct manner in case a uniformly r. e. family is chosen as a hypothesis space, but which are not learnable in the limit, no matter which hypothesis space is chosen. Although such classes have already been offered in the literature, see Angluin [1], up to now all examples—to the authors' knowledge—are defined via diagonalisation in a rather involved manner. In contrast to that, the classes we provide below are very compactly and explicitly defined without any diagonal construction.

Since learning via oracles allows for a characterisation of our query inference type in-between learning in the limit and behaviourally correct learning, we additionally analyse relations between query learning and Gold-style learning for the case that the learners have access to oracles for special non-decidable sets.

This paper summarizes the results published in [16] as well as in [17].

2 Preliminaries and basic results

2.1 Notations

Familiarity with standard mathematical, recursion theoretic, and language theoretic notions and notations is assumed, see Rogers [19] and Hopcroft and Ullman [10]. From now on, a fixed finite alphabet Σ with $\{a, b\} \subseteq \Sigma$ is given, where Σ^* denotes the set of all finite strings over Σ, including the empty string. A *word* is any element from Σ^* and a *language* any subset of Σ^*. The *complement* $\overline{L}$ of a language L is the set $\Sigma^* \setminus L$. $\mathbb{N}$ denotes the set of all natural numbers. Any total function $t : \mathbb{N} \to \Sigma^*$ with $\{t(i) \mid i \in \mathbb{N}\} = L$ is called a *text* for L. A text t is often identified with an infinite sequence $(w_i)_{i\in\mathbb{N}} = (t(i))_{i\in\mathbb{N}}$. Then, for any $n \in \mathbb{N}$, t_n denotes the initial segment $(t(0), \dots, t(n))$ and $content(t_n)$ denotes the set $\{t(0), \dots, t(n)\}$.

In the sequel, let φ be a Gödel numbering of all partial recursive functions and Φ the associated Blum complexity measure, see Blum [5] for a definition. For $i, n \in \mathbb{N}$ we will write $\varphi_i[n]$ for the initial segment $(\varphi_i(0), \dots, \varphi_i(n))$ and say that $\varphi_i[n]$ is defined if all the values $\varphi_i(0), \dots, \varphi_i(n)$ are defined. For convenience, $\varphi_i[-1]$ is always considered defined. Moreover, let $Tot = \{i \in \mathbb{N} \mid \varphi_i \text{ is a total function}\}$ and $K = \{i \in \mathbb{N} \mid \varphi_i(i) \text{ is defined}\}$. The problem to decide for any $i \in \mathbb{N}$ whether or not $\varphi_i(i)$ is defined is called the *halting problem* with respect to φ. As explained by Rogers [19], the halting problem with respect to φ is not decidable and thus the set K is not recursive (neither is the set Tot).

The family $(W_i)_{i\in\mathbb{N}}$ of languages is given by $W_i = \{\omega_j \mid \varphi_i(j) \text{ is defined}\}$ for all $i \in \mathbb{N}$, where $(\omega_j)_{j\in\mathbb{N}}$ is some fixed effective enumeration of Σ^* without repetitions. Moreover, we use a bijective recursive function $\langle \cdot, \cdot \rangle$ coding a pair (x, y) with $x, y \in \mathbb{N}$ into a number $\langle x, y \rangle \in \mathbb{N}$.

If A is any (in general non-recursive) subset of $\mathbb{N}$, then an A-recursive (A-partial recursive) function is a function which is recursive (partial recursive) with the help of an oracle for the set A. That means, an A-recursive (A-partial recursive) function can be computed by an algorithm which has access to an oracle providing correct answers to any question of the type 'does x belong to A?' for $x \in \mathbb{N}$.

A family $(A_i)_{i\in\mathbb{N}}$ of languages is *uniformly recursive* (*uniformly r. e.*) if there is a recursive (partial recursive) function f such that $A_i = \{w \in \Sigma^* \mid f(i, w) = 1\}$

for all $i \in \mathbb{N}$. A family $(A_i)_{i\in\mathbb{N}}$ is *uniformly K-r. e.*, if there is a recursive function g such that $A_i = \{w \in \Sigma^* \mid g(i, w, n) = 1$ for all but finitely many $n\}$ for all $i \in \mathbb{N}$. The notion 'K-r. e.' is related to the notion of A-recursiveness defined above: if $(A_i)_{i\in\mathbb{N}}$ is uniformly K-r. e., this means that there is a K-partial recursive function f such that $A_i = \{w \in \Sigma^* \mid f(i, w) = 1\}$ for all $i \in \mathbb{N}$. Note that for uniformly recursive families membership is uniformly decidable.

Let $\mathcal{C}$ be a class of recursive languages over Σ^*. $\mathcal{C}$ is said to be an *indexable class of recursive languages* (in the sequel we will write *indexable class* for short), if there is a uniformly recursive family $(L_i)_{i\in\mathbb{N}}$ of all and only the languages in $\mathcal{C}$. Such a family will subsequently be called an *indexing* of $\mathcal{C}$.

A family $(T_i)_{i\in\mathbb{N}}$ of *finite* languages is *recursively generable*, if there is a recursive function that, given $i \in \mathbb{N}$, enumerates all elements of T_i and stops.

2.2 Gold-style language learning

Let $\mathcal{C}$ be an indexable class, $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$ any uniformly recursive family (called *hypothesis space*), and $L \in \mathcal{C}$. An *inductive inference machine* (*IIM* for short) M is an algorithmic device that reads longer and longer initial segments σ of a text and outputs numbers $M(\sigma)$ as its hypotheses. An IIM M returning some i is construed to hypothesize the language L_i.

The following definition of learning in the limit is based on Gold [9]. Given a text t for L, *M identifies L from t with respect to $\mathcal{H}$ in the limit*, if

(1) the sequence of hypotheses output by M, when fed t, stabilizes on a number i (i. e., past some point M always outputs the hypothesis i) and
(2) this number i fulfils $L_i = L$.

Moreover, *M identifies L from text with respect to $\mathcal{H}$ in the limit*, if M identifies L from every text with respect to $\mathcal{H}$ in the limit. Finally, *M identifies $\mathcal{C}$ in the limit from text* with respect to $\mathcal{H}$, if it identifies every $L' \in \mathcal{C}$ with respect to $\mathcal{H}$. $\mathit{LimTxt}_{\mathrm{rec}}$ denotes the collection of all indexable classes $\mathcal{C}'$ for which there are an IIM M' and a uniformly recursive family $\mathcal{H}'$ such that M' identifies $\mathcal{C}'$ in the limit from text with respect to $\mathcal{H}'$.

A quite natural and often studied modification of $\mathit{LimTxt}_{\mathrm{rec}}$ is defined by the model of *conservative inference*, see Angluin [1] and Lange and Zeugmann [13]. M is a *conservative* IIM for $\mathcal{C}$ with respect to $\mathcal{H}$, if M performs only justified mind changes, i. e., if M, on some text t for some $L \in \mathcal{C}$, outputs hypotheses i and later j, then M must have seen some element $w \notin L_i$ before returning j. The collection of all indexable classes identifiable in the limit from text by a conservative IIM is denoted by $\mathit{ConsvTxt}_{\mathrm{rec}}$. Note that

$\mathit{ConsvTxt}_{\mathrm{rec}} \subset \mathit{LimTxt}_{\mathrm{rec}}$, as has been shown by Zeugmann and Lange [21].

Since we consider learning from text only, we will assume in the sequel that all languages to be learned are *non-empty*.

One main aspect of human learning is modelled in the approach of learning in the limit: the ability to change one's mind during learning. Thus learning is considered as a process in which the learner may change its hypothesis arbitrarily often until reaching its final correct guess. In particular, it is in general impossible to find out whether or not the final hypothesis has been reached, i. e., whether or not a success in learning has already eventuated.

Another often studied version of Gold-style language learning is behaviourally correct learning, see Case and Lynes [7]: If $\mathcal{C}$ is an indexable class, $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$ a uniformly recursive family, M an IIM, then M is a *behaviourally correct* learner for $\mathcal{C}$ from text with respect to $\mathcal{H}$, if for each $L \in \mathcal{C}$ and each text t for $\mathcal{C}$, all but finitely many outputs i of M when fed t fulfil $L_i = L$. Here M may alternate different correct hypotheses arbitrarily often instead of converging to a single hypothesis. Defining the notion $\mathit{BcTxt}_{\mathrm{rec}}$ correspondingly as usual yields $\mathit{BcTxt}_{\mathrm{rec}} = \mathit{LimTxt}_{\mathrm{rec}}$ (a folklore result). In particular, each IIM *BcTxt*-identifying an indexable class $\mathcal{C}'$ in some uniformly recursive family $\mathcal{H}'$ can be modified to an IIM *LimTxt*-identifying $\mathcal{C}'$ in $\mathcal{H}'$.

This coincidence no longer holds, if more general types of hypothesis spaces are considered. Assume $\mathcal{C}$ is an indexable class and $\mathcal{H}^+$ is any uniformly r. e. family of languages comprising $\mathcal{C}$. Then it is also conceivable to use $\mathcal{H}^+$ as a hypothesis space. $\mathit{LimTxt}_{\mathrm{r.e.}}$ ($\mathit{BcTxt}_{\mathrm{r.e.}}$) denotes the collection of all indexable classes learnable as in the definition of $\mathit{LimTxt}_{\mathrm{rec}}$ ($\mathit{BcTxt}_{\mathrm{rec}}$), if the demand for a uniformly recursive family $\mathcal{H}$ as a hypothesis space is loosened to demanding a uniformly r. e. family $\mathcal{H}^+$ as a hypothesis space. Interestingly, $\mathit{LimTxt}_{\mathrm{rec}} = \mathit{LimTxt}_{\mathrm{r.e.}}$ (a folklore result), i. e., in learning in the limit, the capabilities of IIMs do not increase, if the constraints concerning the hypothesis space are weakened by allowing for arbitrary uniformly r. e. families. A similar relation is obtained for conservative inference; the proof of $\mathit{ConsvTxt}_{\mathrm{rec}} = \mathit{ConsvTxt}_{\mathrm{r.e.}}$ is due to Sanjay Jain (a personal communication). In contrast to that, in the context of *BcTxt*-identification, weakening these constraints yields an add-on in learning power, i. e., $\mathit{BcTxt}_{\mathrm{rec}} \subset \mathit{BcTxt}_{\mathrm{r.e.}}$. In particular, $\mathit{LimTxt}_{\mathrm{rec}} \subset \mathit{BcTxt}_{\mathrm{r.e.}}$ and so *LimTxt*- and *BcTxt*-learning no longer coincide for identification with respect to arbitrary uniformly r. e. families, see also Baliga et al. [4] and Angluin [1].

Hence, in what follows, our analysis of Gold-style language learning will focus on the inference types $\mathit{LimTxt}_{\mathrm{rec}}$, $\mathit{ConsvTxt}_{\mathrm{rec}}$, and $\mathit{BcTxt}_{\mathrm{r.e.}}$. Note that each class in $\mathit{BcTxt}_{\mathrm{r.e.}}$ can also be *BcTxt*-learned with respect to the hypothesis space $(W_i)_{i\in\mathbb{N}}$ (a folklore result, see Gold's results and methods [9]).

The main results of our analysis will be characterisations of these inference types in the query learning model. For that purpose we will make use of well-known characterisations basing on so-called families of *telltales*, see Angluin [1].

Definition 1 *Let $(L_i)_{i\in\mathbb{N}}$ be a uniformly recursive family and $(T_i)_{i\in\mathbb{N}}$ a family of finite non-empty sets. $(T_i)_{i\in\mathbb{N}}$ is called a family of telltales for $(L_i)_{i\in\mathbb{N}}$ iff for all $i, j \in \mathbb{N}$:*

(1) $T_i \subseteq L_i$.
(2) If $T_i \subseteq L_j \subseteq L_i$, then $L_j = L_i$.

The concept of telltale families is the best known notion to illustrate the specific differences between indexable classes in $LimTxt_{\text{rec}}$, $ConsvTxt_{\text{rec}}$, and $BcTxt_{\text{r.e.}}$. Telltale families and their algorithmic structure have turned out to be crucial for learning in our three models, see Angluin [1], Lange and Zeugmann [13], and Baliga et al. [4]:

Theorem 2 *Let $\mathcal{C}$ be an indexable class of languages.*

(1) $\mathcal{C} \in LimTxt_{\text{rec}}$ iff there is an indexing of $\mathcal{C}$ possessing a uniformly r. e. family of telltales.
(2) $\mathcal{C} \in ConsvTxt_{\text{rec}}$ iff there is a uniformly recursive family comprising $\mathcal{C}$ and possessing a recursively generable family of telltales.
(3) $\mathcal{C} \in BcTxt_{\text{r.e.}}$ iff there is an indexing of $\mathcal{C}$ possessing a family of telltales.

The proof of the first statement, see Angluin [1], in particular can be exploited to verify the well-known fact that each indexable class $\mathcal{C} \in LimTxt_{\text{rec}}$ can even be identified in the limit with respect to any indexing $(L_i)_{i\in\mathbb{N}}$ of $\mathcal{C}$. This is not valid correspondingly in the context of conservative inference, as Lange and Zeugmann [13] show.

The notion of telltales is closely related to the notion of *locking sequences*, see Blum and Blum [6]. If $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$ is a hypothesis space, M an IIM, and L a language, then any finite text segment σ of L is called a *LimTxt*-locking sequence (a *BcTxt*-locking sequence) for M, L, and $\mathcal{H}$, if $L_{M(\sigma)} = L$ and $M(\sigma) = M(\sigma\sigma')$ ($L_{M(\sigma)} = L_{M(\sigma\sigma')}$) for all finite text segments σ' of L. If L is *LimTxt*-learned by M (*BcTxt*-learned by M) with respect to $\mathcal{H}$, then there exists a *LimTxt*-locking sequence σ (a *BcTxt*-locking sequence) for M, L, and $\mathcal{H}$. Note that the content of a locking sequence may form a telltale for the corresponding language.

If the requirement '$L_{M(\sigma)} = L$' is dropped, such a sequence σ is called a *stabilizing sequence*, see Fulk [8]. Note that each locking sequence is also a stabilizing sequence, and, if M learns L with respect to $\mathcal{H}$, then each stabilizing sequence for M, L, and $\mathcal{H}$ is a locking sequence for M, L, and $\mathcal{H}$.

In the query learning model, a learner has access to a teacher that truthfully answers queries of a specified kind. A *query learner* M is an algorithmic device that, depending on the reply on the previous queries, either computes a new query or returns a hypothesis and halts, see Angluin [2]. Its queries and hypotheses are coded as natural numbers; both will be interpreted with respect to an underlying *hypothesis space*. When learning an indexable class $\mathcal{C}$, any indexing $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$ of $\mathcal{C}$ may form a hypothesis space. So, as in the original definition, see Angluin [2], when learning $\mathcal{C}$, M is only allowed to query languages belonging to $\mathcal{C}$.

More formally, let $\mathcal{C}$ be an indexable class, let $L \in \mathcal{C}$, let $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$ be an indexing of $\mathcal{C}$, and let M be a query learner. *M learns L with respect to $\mathcal{H}$ using some type of queries* if it eventually halts and its only hypothesis, say i, correctly describes L, i. e., $L_i = L$. So M returns its unique and correct guess i after only finitely many queries. Moreover, *M learns $\mathcal{C}$ with respect to $\mathcal{H}$ using some type of queries*, if it learns every $L' \in \mathcal{C}$ with respect to $\mathcal{H}$ using queries of the specified type. In order to learn a target language L, a query learner M may ask:

Membership queries. The input is an index i. The answer is 'yes' or 'no', depending on whether or not ω_i belongs to L.

Restricted superset queries. The input is an index i. The answer is 'yes' if $L_i \supseteq L$ and 'no' if $L_i \not\supseteq L$.

Restricted disjointness queries. The input is an index i. The answer is 'yes' if $L_i \cap L = \emptyset$ and 'no' if $L_i \cap L \neq \emptyset$.

Note that, except for the case of membership queries, this model in general requires a teacher answering undecidable questions. In particular, asking a restricted superset (restricted disjointness) query with an index i means asking a restricted superset (restricted disjointness) query concerning the language L_i, or, in other words, asking whether or not $L_i \supseteq L$ ($L_i \cap L = \emptyset$).

The term 'restricted' is used to distinguish these types of superset (disjointness) query learning from learning with superset (disjointness) queries, where, together with each negative answer the learner is provided a counterexample, i. e., a word in $L \setminus L_j$ (in $L \cap L_j$). Note that in the current context, it really makes a difference whether or not a query learner is presented counterexamples with each negative reply—both for learning with superset queries and for learning with disjointness queries:

A class learnable with superset queries only in case counterexamples are supplied, can for instance be defined by taking $L_0 = \{a\}^*$ and all languages $L_i = (\{a\}^* \setminus \{a, a^i\}) \cup \{b^z \mid z \geq i\}$, $i \geq 1$, see [15] for further details.

Similarly, choosing $L_0 = \{a^z \mid z \in \mathbb{N}\}$, $L_1 = \{b\}$, and $L_i = \{a^i, b\}$ for $i \geq 2$, one obtains an indexable class which is learnable with disjointness queries just if the learner gets counterexamples with negative replies. To verify this, note that without the help of counterexamples, for all languages L_i with $i \geq 2$, the learning scenarios between the learner and the teacher must be equal and thus the learner has no chance to distinguish these languages from one another. In contrast, if counterexamples are provided, the learner may proceed as follows: First, it poses disjointness queries for L_0 and L_1. In case one of the two queries is answered 'yes', the other of the two queried languages must be the target. In case both are answered 'no', one of the languages L_i with $i \geq 2$ must be the target. But then the counterexample provided with the negative reply to the query for L_0 immediately determines the target language.

MemQ, *rSupQ*, and *rDisQ* denote the collections of all indexable classes $\mathcal{C}'$ for which there are a query learner M' and a hypothesis space $\mathcal{H}'$ such that M' learns $\mathcal{C}'$ with respect to $\mathcal{H}'$ using membership, restricted superset, and restricted disjointness queries, respectively. In the literature, see Angluin [2,3], more types of queries such as (restricted) subset queries and (restricted) equivalence queries have been analysed, but in what follows we concentrate on the three types explained above. Obviously, restricted superset and restricted disjointness queries are in general not decidable, i. e., the teacher may be non-computable.

Note that, in contrast to the models of Gold-style language learning introduced above, learning via queries focusses the aspect of one-shot learning, i. e., it is concerned with learning scenarios in which learning may eventuate without mind changes.

3 Learning with extra queries

Having a closer look at the different models of query learning, one easily finds negative learnability results. For instance, the class $\mathcal{C}_{\text{sup}}$ consisting of the language $L^* = \{a\}^* \cup \{b\}$ and all languages $L^*_i = \{a^k \mid k \leq i\}$, $i \in \mathbb{N}$, is not learnable with restricted superset queries. Assume a query learner M learns $\mathcal{C}_{\text{sup}}$ with restricted superset queries in an indexing $(L_i)_{i\in\mathbb{N}}$ of $\mathcal{C}$ and consider a scenario for M learning L^*. Obviously, a query j is answered 'yes', iff $L_j = L^*$. After finitely many queries, M hypothesizes L^*. Now let i be maximal, such that a query j with $L_j = L^*_i$ has been posed. The above scenario is also feasible for the language L^*_{i+1}. Given this language as a target, M will return a hypothesis representing L^* and thus fail. This yields a contradiction, so $\mathcal{C}_{\text{sup}} \notin rSupQ$.

Moreover, as one can verify easily, the class $\mathcal{C}_{\text{dis}}$ consisting only of the languages

$\{a\}$ and $\{a, b\}$ is not learnable with restricted disjointness queries.

Both examples point to a drawback of Angluin's query model, namely the demand that a query learner is restricted to pose queries concerning languages contained in the class of possible target languages. Note that the class $\mathcal{C}_{\text{sup}}$ would be learnable with restricted superset queries, if it was additionally permitted to query the language $\{a\}^*$, i. e., to ask whether or not this language is a superset of the target language. Similarly, $\mathcal{C}_{\text{dis}}$ would be learnable with restricted disjointness queries, if it was additionally permitted to query the language $\{b\}$. That means there are very simple classes of languages, for which any query learner must fail just because it is barred from asking the 'appropriate' queries.

To overcome this drawback, it seems reasonable to allow the query learner to use an arbitrary (but fixed) uniformly recursive family comprising the target class $\mathcal{C}$ as a hypothesis space in which both its queries and its hypotheses are formulated. So let $\mathcal{C}$ be an indexable class. An *extra query learner* for $\mathcal{C}$ is permitted to query languages in any uniformly recursive family $(L'_i)_{i\in\mathbb{N}}$ comprising $\mathcal{C}$. We say that $\mathcal{C}$ is learnable with extra restricted superset (restricted disjointness) queries with respect to $(L'_i)_{i\in\mathbb{N}}$ iff there is an extra query learner M learning $\mathcal{C}$ with respect to $(L'_i)_{i\in\mathbb{N}}$ using restricted superset (restricted disjointness) queries concerning $(L'_i)_{i\in\mathbb{N}}$. Then $rSupQ_{\text{rec}}$ ($rDisQ_{\text{rec}}$) denotes the collection of all indexable classes $\mathcal{C}$ learnable with extra restricted superset (restricted disjointness) queries with respect to a uniformly recursive family. See also [14,15] for the model of learning with extra queries.

Our classes $\mathcal{C}_{\text{sup}}$ and $\mathcal{C}_{\text{dis}}$ witness $rSupQ \subset rSupQ_{\text{rec}}$ and $rDisQ \subset rDisQ_{\text{rec}}$. Note that both classes would already be learnable, if in addition to the restricted superset (restricted disjointness) queries the learner was allowed to ask a membership query for the word b. So the capabilities of $rSupQ$-learners ($rDisQ$-learners) already increase with the additional permission to ask membership queries. Yet, as Theorem 3 shows, combining restricted superset or restricted disjointness queries with membership queries does not yield the same capability as extra queries do. For convenience, denote the family of classes which are learnable with a combination of restricted superset (restricted disjointness) queries and membership queries by $rSupMemQ$ ($rDisMemQ$).

Theorem 3 *(1)* $rSupQ \subset rSupMemQ \subset rSupQ_{\text{rec}}$.
(2) $rDisQ \subset rDisMemQ \subset rDisQ_{\text{rec}}$.

Proof. We proceed by verifying the following statements:

(A) $rSupQ \subset rSupMemQ$.
(B) $rSupMemQ \subseteq rSupQ_{\text{rec}}$.
(C) $rSupQ_{\text{rec}} \setminus rSupMemQ \neq \emptyset$.
(D) $rDisQ \subset rDisMemQ$.

(E) $rDisMemQ \subseteq rDisQ_{\mathrm{rec}}$.
(F) $rDisQ_{\mathrm{rec}} \setminus rDisMemQ \neq \emptyset$.

Then (A), (B), and (C) together yield (1); (D), (E), and (F) together yield (2).

ad (A).

It is evident that $rSupQ \subseteq rSupMemQ$, whereas the class $\mathcal{C}_{\mathrm{sup}}$ defined above yields an example for a class contained in $rSupMemQ$ but not in $rSupQ$.

ad (B).

In order to prove $rSupMemQ \subseteq rSupQ_{\mathrm{rec}}$, note that, for any word w and any language L, $w \in L$ iff $\Sigma^* \setminus \{w\} \not\supseteq L$. This allows a query learner to simulate membership queries with extra restricted superset queries. Thus, each $rSupMemQ$-learner working with respect to an indexing $(L_i)_{i\in\mathbb{N}}$ can be compiled into an at least equally powerful $rSupQ_{\mathrm{rec}}$-learner which uses an indexing comprising $(L_i)_{i\in\mathbb{N}}$ as well as all languages $\Sigma^* \setminus \{w\}$ for $w \in \Sigma^*$.

ad (C).

$rSupQ_{\mathrm{rec}} \setminus rSupMemQ \neq \emptyset$ is witnessed by the indexable class $\mathcal{C}$ of all languages L_k and $L_{k,j}$ for $k, j \in \mathbb{N}$, where $L_k = \{a^k b^z \mid z \in \mathbb{N}\}$ and

$$L_{k,j} = \begin{cases} L_k\,, & \text{if } \varphi_k(k) \text{ is undefined}\,,\\ \{a^k b^z \mid z \leq \Phi_k(k) \ \vee \ z > \Phi_k(k)+j+1\}\,, & \text{otherwise}\,, \end{cases}$$

see Lange and Zeugmann [13]. Note that $\mathcal{C}$ consists only of infinite languages.

Now (C) is verified in two steps:

(C.1) $\mathcal{C} \in rSupQ_{\mathrm{rec}}$.
(C.2) $\mathcal{C} \notin rSupMemQ$.

ad (C.1).

To verify $\mathcal{C} \in rSupQ_{\mathrm{rec}}$, choose any uniformly recursive family comprising $\mathcal{C}$ and all languages $L_k^* = \{a^k b^z \mid z \leq \Phi_k(k)\}$, $k \in \mathbb{N}$, as a hypothesis space. Note that $L_k^* \in \mathcal{C}$ iff $L_k^* = L_k$ iff $\varphi_k(k)$ is undefined. Let $L \in \mathcal{C}$ be the target language. We define the desired $rSupQ_{\mathrm{rec}}$-learner M for $\mathcal{C}$ as follows. M carries out the instructions below:

1. For $k = 0, 1, 2, \ldots$ ask a restricted superset query concerning L_k, until the answer 'yes' is received for the first time, i. e., until some k with $L_k \supseteq L$ is found.
2. Pose a restricted superset query concerning the language L_k^*.

2.1. If the answer is ‘yes’, then output a hypothesis representing L_k and stop.
2.2. If the answer is ‘no’, then compute $\Phi_k(k)$. Pose a restricted superset query concerning $L_{k,0}$.
   2.2.1. If the answer is ‘no’, then output a hypothesis representing L_k and stop.
   2.2.2. If the answer is ‘yes’, then, for any $j = 1, 2, 3, \ldots$, pose a restricted superset query concerning $L_{k,j}$. As soon as such a query is answered with ‘no’, for some j, output a hypothesis representing $L_{k,j-1}$ and stop.

We prove two claims:

(i) M poses only finitely many queries and then returns a hypothesis;
(ii) this hypothesis correctly describes the target language L.

To prove (i), assume to the contrary that M never returns a hypothesis in a learning process for the target language L. Then either instruction 1 or instruction 2.2.2 (for some k found when carrying out instruction 1) does not terminate.

If instruction 1 does not terminate, then there is no k such that $L_k \supseteq L$. In particular, $L \notin \mathcal{C}$—a contradiction.

If instruction 2.2.2 does not terminate (for the value k found following instruction 1), then $L_k \supseteq L$ and $L_{k,j} \supseteq L$ for all $j \geq 1$. This is only possible if $(L_k^* =)\{a^k b^z \mid z \leq \Phi_k(k)\} \supseteq L$. This again implies $L \notin \mathcal{C}$ and thus a contradiction is deduced. Hence claim (i) follows.

To prove (ii), assume the hypothesis returned by M does not represent L. Then there must be some k found following instruction 1 such that

- M returns a hypothesis h representing L_k in instruction 2.1 or instruction 2.2.1, (a)
- $L_k \neq L$. (b)

By instruction 1, $L_k \supseteq L$. This yields $L_k \supset L$. The latter implies $L = L_{k,j} \neq L_k$ for some j. In particular,

$$\varphi_k(k) \text{ is defined. (c)}$$

If h is returned when executing instruction 2.1, then $L_k^* \supseteq L$. Since $L \in \mathcal{C}$ and $\mathcal{C}$ consists only of infinite languages, L_k^* must be infinite. Hence $\varphi_k(k)$ is undefined, which contradicts (c).

If h is returned when executing instruction 2.2.1, then $L_{k,0} \not\supseteq L$. This implies

$a^k b^{\Phi_k(k)+1} \in L$. The only language in $\mathcal{C}$ for which the latter implies is L_k. Thus $L = L_k$ in contradiction to (b).

By (a), no further cases have to be considered. This establishes claim (ii). Consequently, M learns $\mathcal{C}$ with extra restricted superset queries, which proves (C.1).

ad (C.2).

In contrast to that one can show that $\mathcal{C} \notin rSupMemQ$. Otherwise the halting problem with respect to φ would be decidable, i. e., the set $K = \{i \in \mathbb{N} \mid \varphi_i(i)$ is defined$\}$ would be recursive. To verify this, assume $(L'_i)_{i\in\mathbb{N}}$ is an indexing of $\mathcal{C}$ and M is a query learner which learns $\mathcal{C}$ with membership and restricted superset queries with respect to $(L'_i)_{i\in\mathbb{N}}$. Let $k \geq 0$. A procedure deciding, on input k, whether or not $\varphi_k(k)$ is defined as follows. On input k, the procedure does the following.

- Simulate M.
  Whenever M asks a membership query for some word w, transmit the answer
    'yes' to M, if $w = a^k b^z$ and not $\Phi_k(k) \leq z$,
    'no' to M, if $w \notin \{a^k b^z \mid z \in \mathbb{N}\}$.
  If $w = a^k b^z$ and $z \geq \Phi_k(k)$, then stop and return '1'.
  Whenever M asks a restricted superset query for some language L'_i, transmit the answer
    'yes' to M, if $a^k \in L'_i$ and not $\Phi_k(k) \leq i$,
    'no' to M, if $a^k \notin L'_i$.
  If $a^k \in L'_i$ and $i \geq \Phi_k(k)$, then stop and return '1'.
- In parallel to simulating M start a computation of $\varphi_k(k)$. If $\varphi_k(k)$ is defined before M returns a hypothesis, then stop and return '1'. If M returns a hypothesis before $\varphi_k(k)$ is defined, then stop and return '0'.

It remains to show that this procedure terminates for each $k \in \mathbb{N}$ and that $\varphi_k(k)$ is undefined if the procedure returns '0'.

First, assume there is some $k \in \mathbb{N}$ for which the procedure does not stop. Then $\varphi_k(k)$ is undefined and M does not return any hypothesis in the simulated scenario. (* Note that, in this case, $L_{k,j} = L_k$ for all $j \in \mathbb{N}$ and thus L_k is the only language in $\mathcal{C}$ containing the word a^k. *) But, since all restricted superset queries posed by M correspond to languages in $\mathcal{C}$, the procedure has answered all queries in the simulation of M truthfully concerning the target language $L_k = \{a^k b^z \mid z \in \mathbb{N}\}$. Since $L_k \in \mathcal{C}$, M ought to return a hypothesis in the simulated scenario. This is a contradiction, so the procedure terminates for each $k \in \mathbb{N}$.

Second, assume for some $k \in \mathbb{N}$ that $\varphi_k(k)$ is defined, but the procedure returns '0'. Then M has returned a hypothesis after finitely many queries and answers in the simulated scenario. Now let j be maximal such that a hypothesis or a restricted superset query corresponding to $L_{k,j}$ has been formulated by M.

Consider M when learning $L = L_{k,j+1}$. Obviously, $L_k \supset L$ and $L_{k,i} \supset L$ for all $i < j$. Moreover, none of the membership queries posed by M have concerned words $a^k b^z$ with $z \geq \Phi_k(k)$ (otherwise the procedure would have stopped with the output '1'). Therefore all the queries in the simulated scenario have been answered truthfully with respect to the language L. But by choice of j, the hypothesis returned by M does not represent $L = L_{k,j+1}$. Hence M fails to learn $L_{k,j+1} \in \mathcal{C}$, which contradicts the choice of M. So $\varphi_k(k)$ is undefined whenever the procedure returns '0'.

Thus the set $K = \{i \in \mathbb{N} \mid \varphi_i(i) \text{ is defined}\}$ would be recursive—a contradiction. This implies that $\mathcal{C} \notin rSupMemQ$ and thus (C.2) and (C) are proven.

ad (D).

Obviously, $rDisQ \subseteq rDisMemQ$. Moreover, the class $\mathcal{C}_{\text{dis}}$ defined above witnesses $rDisMemQ \setminus rDisQ \neq \emptyset$.

ad (E).

In order to prove $rDisMemQ \subseteq rDisQ_{\text{rec}}$, note that, for any word w and any language L, $w \in L$ iff $\{w\}$ and L are not disjoint. This allows a query learner to simulate membership queries with extra restricted disjointness queries. Thus, each $rDisMemQ$-learner working with respect to an indexing $(L_i)_{i\in\mathbb{N}}$ can be transformed into an at least equally powerful $rDisQ_{\text{rec}}$-learner which uses an indexing comprising $(L_i)_{i\in\mathbb{N}}$ as well as all singleton languages $\{w\}$ for $w \in \Sigma^*$.

ad (F).

$rDisQ_{\text{rec}} \setminus rDisMemQ \neq \emptyset$ is witnessed by the indexable class $\mathcal{C}$ consisting of $L_0 = \{b\}$ and all languages $L_{i+1} = \{a^{i+1}, b\}$, $i \in \mathbb{N}$.

(F) is verified in two steps:

(F.1) $\mathcal{C} \in rDisQ_{\text{rec}}$.
(F.2) $\mathcal{C} \notin rDisMemQ$.

ad (F.1).

To show that $\mathcal{C} \in rDisQ_{\text{rec}}$, choose a uniformly recursive family comprising $\mathcal{C}$ as well as $\{a\}^*$ and all languages $\{a^{i+1}\}$, $i \in \mathbb{N}$. We define the desired learner M identifying $\mathcal{C}$ with extra restricted disjointness queries as follows. Let M carry out the instructions below.

Pose a restricted disjointness query concerning $\{a\}^*$.
(* Note that the only possible target language disjoint with $\{a\}^*$ is L_0. *)
- If the answer is 'yes', then return a hypothesis representing L_0 and stop.
- If the answer is 'no', then, for $i = 0, 1, 2, \ldots$ ask a restricted disjointness query concerning $\{a^{i+1}\}$, until the answer 'no' is received for the first time. (* Note that this must eventually happen. *)
  As soon as such a query is answered with 'no', for some i, output a hypothesis representing L_{i+1} and stop.

Using the remarks going along with these instructions, it is not hard to verify that M learns $\mathcal{C}$ with extra restricted disjointness queries. This establishes (F.1).

ad (F.2).

In contrast one can show that $\mathcal{C} \notin rDisMemQ$. For that purpose, to deduce a contradiction, assume that there is a query learner M identifying $\mathcal{C}$ with restricted disjointness and membership queries with respect to an indexing $(L'_i)_{i\in\mathbb{N}}$ of $\mathcal{C}$. Consider a learning scenario of M for the target language L_0. Obviously, each restricted disjointness query is answered with 'no'; a membership query for a word w is answered with 'no' iff $w \neq b$. After finitely many queries, M must return a hypothesis representing L_0. Now let i be maximal, such that a membership query concerning a word a^i has been posed. The scenario described above is also feasible for the language $\{a^{i+1}, b\}$. If this language constitutes the target, then M will return a hypothesis representing L_0 and thus fail. This yields the desired contradiction and thus (F.2) and (F) are proven. □

4 New characterisations of Gold-style language learning

One main difference between Gold-style and query learning lies in the question whether or not a current hypothesis of a learner is already correct. A Gold-style learner is allowed to change its mind arbitrarily often (thus in general this question can not be answered), whereas a query learner has to find a correct representation of the target object already in the first guess, i. e., within 'one shot' (and thus the question can always be answered in the affirmative). Another difference is certainly the kind of information provided during the learning process. So, at first glance, these models seem to focus on very different aspects of human learning and do not seem to have much in common.

Thus the question arises, whether there are any similarities in these models at all and whether there are aspects of learning both models capture. Answering this question requires a comparison of both models concerning the capabilities of the corresponding learners. In particular, one central question in this context is whether Gold-style (limit) learners can be replaced by at least equally powerful (one-shot) query learners. The rather trivial examples of classes not learnable with restricted superset or restricted disjointness queries already show that quite general hypothesis spaces—such as in learning with extra queries—are an important demand, if such a replacement shall be successful. In other words, we demand a more potent teacher, able to answer more general questions than in Angluin's original model. Astonishingly, this demand is already sufficient to coincide with the capabilities of conservative limit learners: in [14,15] it has been shown that the collection of indexable classes learnable with extra restricted superset queries coincides with $\mathit{ConsvTxt}_{\mathrm{rec}}$. And, moreover, this also holds for the collection of indexable classes learnable with extra restricted disjointness queries.

Theorem 4 $\mathit{ConsvTxt}_{\mathrm{rec}} = \mathit{rSupQ}_{\mathrm{rec}} = \mathit{rDisQ}_{\mathrm{rec}}$.

Proof. $\mathit{ConsvTxt}_{\mathrm{rec}} = \mathit{rSupQ}_{\mathrm{rec}}$ holds by [14,15]. Thus it remains to prove that $\mathit{rSupQ}_{\mathrm{rec}} = \mathit{rDisQ}_{\mathrm{rec}}$. For that purpose let $\mathcal{C}$ be any indexable class.

First assume $\mathcal{C} \in \mathit{rDisQ}_{\mathrm{rec}}$. Then there is a uniformly recursive family $(L_i)_{i\in\mathbb{N}}$ and a query learner M such that M learns $\mathcal{C}$ with extra restricted disjointness queries with respect to $(L_i)_{i\in\mathbb{N}}$. Now define $L'_{2i} = L_i$ and $L'_{2i+1} = \overline{L_i}$ for all $i \in \mathbb{N}$.

Suppose L is a target language. A query learner M' identifying L with extra restricted superset queries with respect to $(L'_i)_{i\in\mathbb{N}}$ is defined as follows. Let M' execute the following instructions:

- Simulate M when learning L.
- If M poses a restricted disjointness query concerning L_i, then pose a restricted superset query concerning L'_{2i+1} to your teacher. If the answer is 'yes', then transmit the answer 'yes' to M. If the answer is 'no', then transmit the answer 'no' to M.
  (* Note that $L_i \cap L = \emptyset$ iff $\overline{L_i} \supseteq L$ iff $L'_{2i+1} \supseteq L$. *)
- If M hypothesizes L_i, then output a representation for L'_{2i}.

It is not hard to verify that M' learns $\mathcal{C}$ with extra restricted superset queries with respect to $(L'_i)_{i\in\mathbb{N}}$. Hence $\mathcal{C}$ belongs to $\mathit{rSupQ}_{\mathrm{rec}}$, which finally implies $\mathit{rDisQ}_{\mathrm{rec}} \subseteq \mathit{rSupQ}_{\mathrm{rec}}$.

The opposite inclusion $\mathit{rSupQ}_{\mathrm{rec}} \subseteq \mathit{rDisQ}_{\mathrm{rec}}$ is verified analogously. □

Note that this proof is mainly based on the fact that the complement of any recursive language is recursive itself. Thus restricted superset queries can be used instead of restricted disjointness queries, if the query and hypothesis space may be an arbitrary uniformly recursive family.

As initially in Gold-style learning, we have only considered uniformly recursive families as hypothesis spaces for query learners so far. Similarly to the notion of $BcTxt_{\text{r.e.}}$, it is conceivable to permit more general hypothesis spaces also in the query model, i.e., to demand an even more capable teacher. Thus, by $rSupQ_{\text{r.e.}}$ ($rDisQ_{\text{r.e.}}$) we denote the collection of all indexable classes which are learnable with restricted superset (restricted disjointness) queries with respect to a uniformly r.e. family. Similarly to Gold-style learning with respect to uniformly r.e. hypothesis spaces, note that each class in $rSupQ_{\text{r.e.}}$ ($rDisQ_{\text{r.e.}}$) can also be learned with restricted superset (restricted disjointness) queries with respect to our fixed Gödel numbering $(W_i)_{i\in\mathbb{N}}$.

Interestingly, this relaxation helps to characterise learning in the limit in terms of query learning.

Theorem 5 $LimTxt_{\text{rec}} = rDisQ_{\text{r.e.}}$.

Proof. First, we show $rDisQ_{\text{r.e.}} \subseteq LimTxt_{\text{rec}}$. For that purpose, let $\mathcal{C} \in rDisQ_{\text{r.e.}}$ be an indexable class. Fix a query learner M identifying $\mathcal{C}$ with restricted disjointness queries with respect to $(W_i)_{i\in\mathbb{N}}$.

The IIM M', which is defined as follows, $LimTxt$-identifies $\mathcal{C}$ with respect to $(W_i)_{i\in\mathbb{N}}$. Given an initial segment t_n of some text t, M' interacts with M simulating a learning process for n steps.

In step k, $k \leq n$, depending on how M' has replied to the previous queries posed by M, the learner M computes either (i) a new query i or (ii) a hypothesis i.

- In case (ii), M' returns the hypothesis i and stops simulating M.
- In case (i), M' checks whether there is a word in $content(t_n)$, which is found in W_i within n steps. If such a word exists, M' transmits the answer 'no' to M; else M' transmits the answer 'yes' to M. If $k < n$, then M executes step $k+1$, else M' returns any auxiliary hypothesis and stops simulating M.

Given segments $t_n, t_{n+1}, t_{n+2}, \ldots$ of a text t for some target language, if n is large enough, M' answers all queries of M correctly and M returns its sole hypothesis within n steps. So the hypotheses returned by M' stabilize on this correct guess.

Hence $\mathcal{C} \in LimTxt_{\text{r.e.}} (= LimTxt_{\text{rec}})$ and therefore $rDisQ_{\text{r.e.}} \subseteq LimTxt_{\text{rec}}$.

Second, we show that $LimTxt_{\text{rec}} \subseteq rDisQ_{\text{r.e.}}$. So let $\mathcal{C} \in LimTxt_{\text{rec}}$ be an indexable class. Fix an indexing $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$ of $\mathcal{C}$ and an IIM M, such that

M *LimTxt*-identifies $\mathcal{C}$ with respect to $\mathcal{H}$ (recall the remark below Theorem 2 to see that suchlike $\mathcal{H}$ and M exist).

Suppose $L \in \mathcal{C}$ is the target language. An *rDisQ*-learner M' for L with respect to $(W_i)_{i\in\mathbb{N}}$ is defined as follows. M' executes the instructions below, starting in step 0. Note that Gödel numbers (representations in $(W_i)_{i\in\mathbb{N}}$) can be computed for all queries to be asked. Step n reads as follows:

- Ask restricted disjointness queries for $\{\omega_0\}, \ldots, \{\omega_n\}$. Let $L_{[n]}$ be the set of words ω_x, $x \leq n$, for which the corresponding query is answered with 'no'. (* Note that $L_{[n]} = L \cap \{\omega_x \mid x \leq n\}$. *)
- Let $(\sigma_x^n)_{x\in\mathbb{N}}$ be an effective enumeration of all finite text segments for $L_{[n]}$. For all $x, y \leq n$ pose a restricted disjointness query for $\overline{L_{M(\sigma_x^y)}}$ and thus build $\mathrm{Cand}_n = \{\sigma_x^y \mid x, y \leq n \text{ and } \overline{L_{M(\sigma_x^y)}} \cap L = \emptyset\}$ from the queries answered with 'yes'.
  (* Note that $\mathrm{Cand}_n = \{\sigma_x^y \mid x, y \leq n \text{ and } L \subseteq L_{M(\sigma_x^y)}\}$. *)
- For all $\sigma \in \mathrm{Cand}_n$, pose a restricted disjointness query for the language

$$W'_\sigma = \begin{cases} \Sigma^*, & \text{if } M(\sigma\sigma') \neq M(\sigma) \text{ for some text segment } \sigma' \text{ of } L_{M(\sigma)}\,, \\ \emptyset, & \text{otherwise}\,. \end{cases}$$

  (* Note that W'_σ is uniformly r. e. in σ and $W'_\sigma \cap L = \emptyset$ iff σ is a *LimTxt*-stabilizing sequence for M and $L_{M(\sigma)}$. *)
  If all these restricted disjointness queries are answered with 'no', then go to step $n + 1$.
  Otherwise, if $\sigma \in \mathrm{Cand}_n$ is minimal fulfilling $W'_\sigma \cap L = \emptyset$, then return a hypothesis representing $L_{M(\sigma)}$ and stop.

M' identifies L with restricted disjointness queries with respect to $(W_i)_{i\in\mathbb{N}}$, because (i) M' eventually returns a hypothesis and (ii) this hypothesis is correct for L. To prove (i), note that M is a *LimTxt*-learner for L with respect to $(L_i)_{i\in\mathbb{N}}$. So there are i, x, y such that $M(\sigma_x^y) = i$, $L_i = L$, and σ_x^y is a *LimTxt*-locking sequence for M and L. Then $W'_{\sigma_x^y} = \emptyset$ and the corresponding restricted disjointness query is answered with 'yes'. Thus M' returns a hypothesis. To prove (ii), assume M' returns a hypothesis representing $L_{M(\sigma)}$ for some text segment σ of L. Then, by definition of M', $L \subseteq L_{M(\sigma)}$ and σ is a *LimTxt*-stabilizing sequence for M and $L_{M(\sigma)}$. In particular, σ is a *LimTxt*-stabilizing sequence for M and L. Since M learns L in the limit from text with respect to $\mathcal{H} = (L_i)_{i\in\mathbb{N}}$, this implies that σ is even a *LimTxt*-locking sequence for M, L and $\mathcal{H}$, and thus $L = L_{M(\sigma)}$. Hence the hypothesis M' returns is correct for L.

Therefore $\mathcal{C} \in rDisQ_{\text{r.e.}}$ and $LimTxt_{\text{rec}} \subseteq rDisQ_{\text{r.e.}}$. □

Comparing this result to Theorem 4, it is evident that a characterisation of $LimTxt_{\text{rec}}$ in terms of learning with restricted superset queries is missing.

Indeed, if one tries to adopt the proof of $rDisQ_{\text{rec}} = rSupQ_{\text{rec}}$ for the general case of learning with respect to uniformly r. e. families, the obstacle is the simple fact that the complement of an r. e. language is not necessarily r. e. itself. So, whereas in uniformly recursive families each restricted superset query for some language can be simulated by a restricted disjointness query for its complement (and vice versa), this is no longer valid in the context of uniformly r. e. query and hypothesis spaces. Thus there remains the question whether or not $rDisQ_{\text{r.e.}}$ equals $rSupQ_{\text{r.e.}}$. We shall address this point again in Section 5.

Reducing the constraints concerning the hypothesis spaces even more, let $rSupQ_{K\text{-r.e.}}$ ($rDisQ_{K\text{-r.e.}}$) denote the collection of all indexable classes which are learnable using restricted superset (restricted disjointness) queries with respect to a uniformly K-r. e. family. With analogous definitions for Gold-style learning one easily obtains $LimTxt_{K\text{-r.e.}} = LimTxt_{\text{r.e.}} = LimTxt_{\text{rec}}$ and $BcTxt_{K\text{-r.e.}} = BcTxt_{\text{r.e.}}$. This finally allows for a characterisation of the classes in $BcTxt_{\text{r.e.}}$.

Theorem 6 $BcTxt_{\text{r.e.}} = rSupQ_{K\text{-r.e.}} = rDisQ_{K\text{-r.e.}}$.

Proof. First we show $rSupQ_{K\text{-r.e.}} \subseteq BcTxt_{\text{r.e.}}$ and $rDisQ_{K\text{-r.e.}} \subseteq BcTxt_{\text{r.e.}}$. For that purpose, let $\mathcal{C} \in rSupQ_{K\text{-r.e.}}$ ($\mathcal{C} \in rDisQ_{K\text{-r.e.}}$) be an indexable class, $(L_i)_{i\in\mathbb{N}}$ an indexing of $\mathcal{C}$. Fix a uniformly K-r. e. family $(V_i)_{i\in\mathbb{N}}$ and a query learner M identifying $\mathcal{C}$ with restricted superset (restricted disjointness) queries with respect to $(V_i)_{i\in\mathbb{N}}$.

To obtain a contradiction, assume that $\mathcal{C} \notin BcTxt_{\text{r.e.}}$. By Theorem 2, $(L_i)_{i\in\mathbb{N}}$ does not possess a telltale family. In other words, there is some $i \in \mathbb{N}$, such that for any finite set $W \subseteq L_i$ there exists some $j \in \mathbb{N}$ satisfying $W \subseteq L_j \subset L_i$. $(*)$

Consider M when learning L_i. In the corresponding learning scenario S the learner M poses finitely many queries and afterwards returns a hypothesis representing L_i. Let

- $V_{i_1^-}, \ldots, V_{i_k^-}$ be the languages for which M poses queries in the scenario S which are answered 'no';
- $V_{i_1^+}, \ldots, V_{i_m^+}$ be the languages for which M poses queries in the scenario S which are answered 'yes'.

That means, for all $z \in \{1, \ldots, k\}$, we have $V_{i_z^-} \not\supseteq L_i$ ($V_{i_z^-} \cap L_i \neq \emptyset$). In particular, for all $z \in \{1, \ldots, k\}$, there is a word $w_z \in L_i \setminus V_{i_z^-}$ ($w_z \in V_{i_z^-} \cap L_i$). Let $W = \{w_1, \ldots, w_k\}(\subseteq L_i)$. By $(*)$ there is some $j \in \mathbb{N}$ satisfying $W \subseteq L_j \subset L_i$.

Now note that the above scenario S is also feasible for L_j: $w_z \in L_j$ implies $V_{i_z^-} \not\supseteq L_j$ ($V_{i_z^-} \cap L_j \neq \emptyset$) for all $z \in \{1, \ldots, k\}$. $V_{i_z^+} \supseteq L_i$ ($V_{i_z^+} \cap L_i = \emptyset$) implies $V_{i_z^+} \supseteq L_j$ ($V_{i_z^+} \cap L_j = \emptyset$) for all $z \in \{1, \ldots, m\}$. Thus all queries in S are answered truthfully for L_j. Since M hypothesizes L_i in the scenario S, and $L_i \neq L_j$, M fails to identify L_j. This is the desired contradiction.

Hence $\mathcal{C} \in \mathit{BcTxt}_{\text{r.e.}}$, so $\mathit{rSupQ}_{K\text{-r.e.}} \subseteq \mathit{BcTxt}_{\text{r.e.}}$, $\mathit{rDisQ}_{K\text{-r.e.}} \subseteq \mathit{BcTxt}_{\text{r.e.}}$.

Second we show that $\mathit{BcTxt}_{\text{r.e.}} \subseteq \mathit{rSupQ}_{K\text{-r.e.}}$ and $\mathit{BcTxt}_{\text{r.e.}} \subseteq \mathit{rDisQ}_{K\text{-r.e.}}$. So let $\mathcal{C} \in \mathit{BcTxt}_{\text{r.e.}}$ be an indexable class. Fix an IIM M which learns $\mathcal{C}$ according to the definition of $\mathit{BcTxt}_{\text{r.e.}}$ with respect to $(W_i)_{i\in\mathbb{N}}$.

Let $(V_i)_{i\in\mathbb{N}}$ be a uniformly K-r. e. family such that indices can be computed for all queries to be asked below.

Assume $L \in \mathcal{C}$ is the target language. A query learner M' identifying L with restricted superset (restricted disjointness) queries with respect to $(V_i)_{i\in\mathbb{N}}$ is defined as follows. Let M carry out the instructions below, starting in step 0. Step n reads as follows:

- Ask restricted superset queries for $\Sigma^* \setminus \{\omega_i\}$ (restricted disjointness queries for $\{\omega_i\}$) for all $i \leq n$. Let $L_{[n]}$ be the set of words ω_x, $x \leq n$, for which the corresponding query is answered with 'no'.
  (* Note that $L_{[n]} = L \cap \{\omega_x \mid x \leq n\}$. *)
- Let $(\sigma_x^n)_{x\in\mathbb{N}}$ be an effective enumeration of all finite text segments for $L_{[n]}$. For all $x, y \leq n$ pose a restricted superset query for $W_{M(\sigma_x^y)}$ (a restricted disjointness query for $\overline{W_{M(\sigma_x^y)}}$) and thus build $\text{Cand}_n = \{\sigma_x^y \mid x, y \leq n \text{ and } W_{M(\sigma_x^y)} \supseteq L\} = \{\sigma_x^y \mid x, y \leq n \text{ and } \overline{W_{M(\sigma_x^y)}} \cap L = \emptyset\}$ from the queries answered with 'yes'.
- For all $\sigma \in \text{Cand}_n$, pose a restricted superset (restricted disjointness) query for the language

$$V'_\sigma = \begin{cases} \Sigma^*, & \text{if } W_{M(\sigma)} \neq W_{M(\sigma\sigma')} \text{ for some text segment } \sigma' \text{ of } W_{M(\sigma)}\,, \\ \emptyset, & \text{otherwise}\,. \end{cases}$$

  (* Note that V'_σ is uniformly K-r. e. in σ and $V'_\sigma \not\supseteq L$ iff $V'_\sigma \cap L = \emptyset$ iff σ is a BcTxt-stabilizing sequence for M, $W_{M(\sigma)}$, and $(W_i)_{i\in\mathbb{N}}$. *)
  If all these restricted superset queries are answered with 'yes' (all these restricted disjointness queries are answered with 'no'), then go to step $n+1$. Otherwise, if $\sigma \in \text{Cand}_n$ is minimal fulfilling $V'_\sigma \not\supseteq L$ and thus $V'_\sigma \cap L = \emptyset$, then return a hypothesis representing $W_{M(\sigma)}$ and stop.

M' learns L with restricted superset (restricted disjointness) queries in $(V_i)_{i\in\mathbb{N}}$, because (i) M' eventually returns a hypothesis and (ii) this hypothesis is correct for L. To prove (i), note that M is a BcTxt-learner for L in $(W_i)_{i\in\mathbb{N}}$. So there are x, y such that $W_{M(\sigma_x^y)} = L$ and σ_x^y is a BcTxt-locking sequence for M, L, and $(W_i)_{i\in\mathbb{N}}$. Then $V'_{\sigma_x^y} = \emptyset$ and the corresponding restricted superset query is answered with 'no' (the restricted disjointness query with 'yes'). Thus

M' returns a hypothesis. To prove (ii), suppose M' returns a hypothesis representing $W_{M(\sigma)}$ for a text segment σ of L. Then, by definition of M', σ is a $BcTxt$-stabilizing sequence for M, $W_{M(\sigma)}$, and $(W_i)_{i\in\mathbb{N}}$. In particular, σ is a $BcTxt$-stabilizing sequence for M, L, and $(W_i)_{i\in\mathbb{N}}$. As M $BcTxt$-learns L, this implies that σ is a $BcTxt$-locking sequence for M, L, and $(W_i)_{i\in\mathbb{N}}$. Hence $L = W_{M(\sigma)}$ and the hypothesis of M' is correct for L.

Therefore $\mathcal{C} \in rSupQ_{K\text{-r.e.}} \cap rDisQ_{K\text{-r.e.}}$, and thus $BcTxt_{\text{r.e.}} \subseteq rSupQ_{K\text{-r.e.}}$ and $BcTxt_{\text{r.e.}} \subseteq rDisQ_{K\text{-r.e.}}$. □

Note that, although the complement of a K-r. e. language is not necessarily K-r. e. itself, we obtain $rDisQ_{K\text{-r.e.}} = rSupQ_{K\text{-r.e.}}$. In particular, the latter equivalence cannot be verified with the method used for proving $rDisQ_{\text{rec}} = rSupQ_{\text{rec}}$ (see the proof of Theorem 4 and the proximate remark).

5 The grading of restricted superset queries

The characterisations verified in the preceding section have revealed a correspondence between Gold-style learning and learning via queries—between limiting and one-shot learning processes.

Crucial in this context is that the learner may ask the 'appropriate' queries. Thus the choice of hypothesis spaces and, correspondingly, the ability of the teacher is decisive. If the teacher is capable of answering restricted disjointness queries in some uniformly r. e. family of languages, then, by Theorem 5, learning with restricted disjointness queries coincides with learning in the limit. Interestingly, given uniformly recursive or uniformly K-r. e. families as hypothesis spaces, restricted disjointness and restricted superset queries can be considered equally powerful tools for query learners, i. e., the capabilities of learners using restricted disjointness queries are equal to those of learners using restricted superset queries. As it turns out, this relation is not valid, if the hypothesis space may be any uniformly r. e. family. That means, $rDisQ_{\text{r.e.}}$ (and $LimTxt_{\text{rec}}$) is not equal to the collection of all indexable classes learnable with restricted superset queries in uniformly r. e. families.

Theorem 7 $LimTxt_{\text{rec}} \subset rSupQ_{\text{r.e.}}$.

Proof. To verify $LimTxt_{\text{rec}} \subseteq rSupQ_{\text{r.e.}}$, the proof of $LimTxt_{\text{rec}} \subseteq rDisQ_{\text{r.e.}}$ can be adapted. It remains to quote a class in $rSupQ_{\text{r.e.}} \setminus LimTxt_{\text{rec}}$.

Let, for all $k, j \in \mathbb{N}$, $\mathcal{C}_{\text{lim}}$ contain the languages $L_k = \{a^k b^z \mid z \geq 0\}$ and

$$L_{k,j} = \begin{cases} \{a^k b^z \mid z \leq m\}\,, & \text{if } m \leq j \text{ is minimal such} \\ & \quad \text{that } \varphi_k(m) \text{ is undefined}\,, \\ \{a^k b^z \mid z \leq j\} \cup \{b^{j+1} a^{y+1}\}\,, & \text{if } \varphi_k(x) \text{ is defined for all } x \leq j \\ & \quad \text{and } y = \max\{\Phi_k(x) \mid x \leq j\}\,. \end{cases}$$

$\mathcal{C}_{\text{lim}}$ is an indexable class; the proof is omitted.

To show $\mathcal{C}_{\text{lim}} \in rSupQ_{\text{r.e.}}$, assume $L \in \mathcal{C}_{\text{lim}}$ is the target language. A learner M identifying L with restricted superset queries with respect to $(W_i)_{i\in\mathbb{N}}$ is defined as follows. Let M execute the following instructions:

- For $k = 0, 1, 2, \ldots$ ask a restricted superset query concerning $L_k \cup \{b^r a^s \mid r, s \in \mathbb{N}\}$, until the answer 'yes' is received for the first time.
- Pose a restricted superset query concerning the language L_k.
  If the answer is 'no', then, for $r, s = 0, 1, 2, \ldots$ ask a restricted superset query concerning $L_k \cup \{b^{r+1} a^{s+1}\}$, until the answer 'yes' is received for the first time. Output a hypothesis representing $L_{k,r}$ and stop.
  If the answer is 'yes', then pose a restricted superset query for the language

$$W'_k = \begin{cases} \{a^k b^z \mid z \leq j\}\,, & \text{if } j \text{ is minimal, such that } \varphi_k(j) \text{ is undefined}\,, \\ \{a^k b^z \mid z \geq 0\}\,, & \text{if } \varphi_k \text{ is a total function}\,. \end{cases}$$

  (* Note that W'_k is uniformly r. e. in k. W'_k is a superset of L iff $W'_k = L$. *)
  If the answer is 'yes', then return a hypothesis representing W'_k and stop.
  If the answer is 'no', then return a hypothesis representing L_k and stop.

The remarks going along with these instructions show that M is an *rSupQ*-learner for $\mathcal{C}_{\text{lim}}$ with respect to $(W_i)_{i\in\mathbb{N}}$.

Finally, $\mathcal{C}_{\text{lim}} \notin \mathit{LimTxt}_{\text{rec}}$ holds, since otherwise *Tot* would be K-recursive. To verify this, assume M is an IIM learning $\mathcal{C}_{\text{lim}}$ in the limit from text. A procedure deciding, for any $k \geq 0$, whether or not φ_k is a total function, is defined as follows:

- Let σ be a *LimTxt*-locking sequence for M and L_k.
  (* Note that σ exists by assumption and thus can be found by a K-recursive procedure. *)
- If there is some $x \leq \max\{z \mid a^k b^z \text{ occurs in } \sigma\}$, such that $\varphi_k(x)$ is undefined (* also a K-recursive test *), then return '0'. Otherwise return '1'.

It remains to show that φ_k is total, if this procedure returns '1'. So let the procedure return '1'. Assume φ_k is not total and j is minimal, such that $\varphi_k(j)$

is undefined. By definition, the language $L = \{a^k b^z \mid z \leq j\}$ belongs to $\mathcal{C}_{\text{lim}}$. Then the sequence σ found in the procedure is also a text segment for L and by choice—since $L \subset L_k$—a *LimTxt*-locking sequence for M and L. As $M(\sigma)$ is correct for L_k, M fails to identify L. This is a contradiction; hence φ_k is total.

Thus the set *Tot* is K-recursive—a contradiction. So $\mathcal{C}_{\text{lim}} \notin \mathit{LimTxt}_{\text{rec}}$. □

Since $\mathit{rSupQ}_{\text{r.e.}} \subseteq \mathit{rSupQ}_{K\text{-r.e.}}$, one easily obtains $\mathit{rSupQ}_{\text{r.e.}} \subseteq \mathit{BcTxt}_{\text{r.e.}}$ from Theorem 6.

The more challenging question is whether or not the inference types $\mathit{rSupQ}_{\text{r.e.}}$ and $\mathit{rSupQ}_{K\text{-r.e.}}$ coincide and thus whether or not $\mathit{rSupQ}_{\text{r.e.}}$ equals $\mathit{BcTxt}_{\text{r.e.}}$. Interestingly, this is not the case, that means, when compared to uniformly r. e. families, K-r. e. numberings provide a further benefit for learning with restricted superset queries.

Theorem 8 $\mathit{rSupQ}_{\text{r.e.}} \subset \mathit{rSupQ}_{K\text{-r.e.}}$.

Though our current tools allow for a verification of this theorem, the proof would be rather lengthy. Since a characterisation of $\mathit{rSupQ}_{\text{r.e.}}$ in terms of Gold-style learning simplifies the proof considerably, we postpone the proof for now.

By Theorems 7 and 6 the statement of Theorem 8 implies $\mathit{LimTxt}_{\text{rec}} \subset \mathit{rSupQ}_{\text{r.e.}} \subset \mathit{BcTxt}_{\text{r.e.}}$, i. e., we have found a natural type of learners the capabilities of which are strictly between those of *LimTxt*-learners and those of *BcTxt*-learners. This raises the question whether the learning type $\mathit{rSupQ}_{\text{r.e.}}$ can also be characterised in terms of Gold-style learning. This is indeed possible if we consider learners which have access to some oracle and may use uniformly r. e. numberings as their hypothesis spaces. In the sequel the notion $\mathit{ConsvTxt}_{\text{r.e.}}[K]$ refers to the collection of indexable classes which are learnable in the sense of $\mathit{ConsvTxt}_{\text{rec}}$, if (i) K-recursive IIMs are considered as learners and (ii) uniformly r. e. numberings are admitted as hypothesis spaces. For more background on learning with oracles, the reader is directed to Stephan [20].

Theorem 9 $\mathit{rSupQ}_{\text{r.e.}} = \mathit{ConsvTxt}_{\text{r.e.}}[K]$.

Proof. First, we prove $\mathit{rSupQ}_{\text{r.e.}} \subseteq \mathit{ConsvTxt}_{\text{r.e.}}[K]$. For that purpose assume $\mathcal{C}$ is an indexable class in $\mathit{rSupQ}_{\text{r.e.}}$. Let M be a query learner identifying $\mathcal{C}$ in $(W_i)_{i\in\mathbb{N}}$ and assume without loss of generality that each hypothesis ever returned by M corresponds to the intersection of all queries answered with 'yes' in the preceding scenario.
(* Think of M as a normalisation of a restricted superset query learner M^-: M copies M^- until M^- returns the hypothesis i. Now M asks a query for the language W_i instead of returning a hypothesis. Then let M return a hypothesis

j representing the intersection of all queries answered with 'yes' in its preceding scenario. Given a fair scenario for W_i and a successful learner M^-, this implies $W_i = W_j$ and thus M is successful. *)

Let $L \in \mathcal{C}$, t a text for L. A conservative learner M' is defined as follows. Let $M'(t_0)$ be an index of the language *content*(t_0). On input t_n for $n \geq 1$, M' computes $M'(t_n)$ following the procedure below:

M' simulates M for n steps of computation.
- Whenever M asks a restricted superset query i, M' transmits the answer 'yes' to M, if *content*$(t_n) \subseteq W_i$, the answer 'no', otherwise.
(* This test is K-recursive. *)
- If M returns a hypothesis i within n steps of computation, let M' return i on t_n; otherwise let $M'(t_n) = M'(t_{n-1})$.

Note that there must be some n, such that M' answers all queries of M truthfully respecting L. Thus it is not hard to verify that the K-recursive IIM M' learns L in the limit from text. Moreover, $W_{M'(t_n)} \not\supset L$ for all n: assuming $W_{M'(t_n)} \supset L$ implies, by normalisation of M, that all queries M' has answered with 'yes' in the simulation of M indeed represent supersets of L. Since all 'no'-answers are truthful respecting L by definition, this yields a valid query-scenario for L. As M learns L from restricted superset queries, the hypothesis i must correctly describe L—a contradiction. So M' learns $\mathcal{C}$ without ever returning an index of a proper superset of a language currently to be identified. Now it is not hard to modify M' into a K-recursive IIM which works conservatively for the class $\mathcal{C}$ (a hypothesis will only be changed if its inconsistency is verified with the help of a K-oracle). Thus $\mathcal{C} \in \mathit{ConsvTxt}_{\text{r.e.}}[K]$ and $\mathit{rSupQ}_{\text{r.e.}} \subseteq \mathit{ConsvTxt}_{\text{r.e.}}[K]$.

Second, we show $\mathit{ConsvTxt}_{\text{r.e.}}[K] \subseteq \mathit{rSupQ}_{\text{r.e.}}$. For that purpose assume $\mathcal{C}$ is an indexable class in $\mathit{ConsvTxt}_{\text{r.e.}}[K]$. Let M be a K-recursive IIM identifying $\mathcal{C}$ with respect to $(W_i)_{i \in \mathbb{N}}$. Suppose $L \in \mathcal{C}$ is the target language. An *rSupQ*-learner M' for L with respect to $(W_i)_{i \in \mathbb{N}}$ is defined by steps as follows, starting in step 0. Note that representations in $(W_i)_{i \in \mathbb{N}}$ can be computed for all queries to be asked. In step 0, M' finds the minimal m, such that the query for $\Sigma^* \setminus \{\omega_m\}$ is answered 'no'. M' sets $t(0) = \omega_m$ and goes to step 1. In general, step $n+1$ reads as follows:

- Ask a restricted superset query for $\Sigma^* \setminus \{\omega_{n+1}\}$. If the answer is 'no', let $t(n+1) = \omega_{n+1}$; if the answer is 'yes', let $t(n+1) = t(n)$.
(* Note that *content*$(t_{n+1}) = L \cap \{\omega_x \mid x \leq n+1\}$. *)
- Simulate M on input t_{n+1}. Whenever M wants to access a K-oracle for the question whether $j \in K$, formulate a restricted superset query for the

language

$$W'_j = \begin{cases} \Sigma^*, & \text{if } \varphi_j(j) \text{ is defined}, \\ \emptyset, & \text{otherwise}. \end{cases}$$

and transmit the received answer to M.
(* Note that W'_j is uniformly r. e. in j and $W'_j \supseteq L$ iff $\varphi_j(j)$ is defined. *)
As soon as M returns $i = M(t_{n+1})$, pose a restricted superset query for W_i.
If the answer is 'yes', then return the hypothesis i and stop.
(* Since M learns L conservatively, we have $W_i \not\supset L$ and thus $W_i = L$. *)
If the answer is 'no', then go to step $n + 2$.

Using the remarks going along with these instructions, it is not hard to verify that M' learns $\mathcal{C}$ with restricted superset queries in $(W_i)_{i\in\mathbb{N}}$. Thus $\mathcal{C} \in rSupQ_{\text{r.e.}}$ and $ConsvTxt_{\text{r.e.}}[K] \subseteq rSupQ_{\text{r.e.}}$. □

Applying this characterisation, Theorem 8 translates as follows:

Theorem 10 $ConsvTxt_{\text{r.e.}}[K] \subset BcTxt_{\text{r.e.}}$.

Proof. By Theorems 6 and 9 it suffices to prove $BcTxt_{\text{r.e.}} \not\subseteq ConsvTxt_{\text{r.e.}}[K]$. For that purpose we provide an indexable class $\mathcal{C}_{\text{bc}}$ which is $BcTxt$-learnable with respect to some uniformly r. e. numbering, but not learnable according to the definition of $ConsvTxt_{\text{r.e.}}[K]$. For all $k \in \mathbb{N}$, $\mathcal{C}_{\text{bc}}$ contains the language $L_k = \{a^k b^z \mid z \geq 0\}$. Moreover, for all $k, i, j \in \mathbb{N}$ for which $\varphi_k[i-1]$ is defined and $j \leq i$, let $\mathcal{C}_{\text{bc}}$ contain the language

$$L_{k,i,j} = \begin{cases} \{a^k b^z \mid z \leq j\}, & \text{if } \varphi_k(i) \text{ is undefined}, \\ \{a^k b^z \mid z \leq j\} \cup \{ba^{\Phi_k(i)}\}, & \text{if } \varphi_k(i) \text{ is defined}. \end{cases}$$

To show that $\mathcal{C}_{\text{bc}} \in BcTxt$, it suffices by Theorem 2 to prove the existence of telltales corresponding to some indexing of $\mathcal{C}_{\text{bc}}$. This is quite simple: as each language $L_{k,i,j} \in \mathcal{C}_{\text{bc}}$ is finite, it forms a telltale for itself. Moreover, as for all k there are only finitely many subsets of L_k in $\mathcal{C}_{\text{bc}}$, telltales for L_k must exist, too.

Finally, it remains to prove that $\mathcal{C}_{\text{bc}} \notin ConsvTxt_{\text{r.e.}}[K]$. Assume the opposite, i. e., there is some K-recursive IIM M which $ConsvTxt$-identifies $\mathcal{C}_{\text{bc}}$ in $(W_i)_{i\in\mathbb{N}}$. The idea is to deduce a contradiction by concluding that Tot is K-recursive. For that purpose, define a K-recursive procedure on input k as follows:

- Let $t = a^k, a^k b, a^k b^2, \ldots$ be the 'canonical' text for L_k.
- Simulate M on input $t_0, t_1, t_2, \ldots$ until some n is found with $content(t_n) \subset W_{M(t_n)} \subseteq L_k$.
  (* n exists, as M learns L_k. Determining n is K-recursive. *)

- If $\varphi_k(i)$ is defined for all $i \leq n$, then return '1'; otherwise return '0'.

Obviously, this procedure is K-recursive. Note that it returns '0' only in case φ_k is not total. So assume it returns '1'. Then there is some n such that $content(t_n) \subset W_{M(t_n)} \subseteq L_k$. If φ_k was not total, the minimal i for which $\varphi_k(i)$ is undefined would be greater than n. Thus $L = \{a^k b^z \mid z \leq n\} \in \mathcal{C}_{\mathrm{bc}}$. Now t_n is also a text segment for L, but $L = content(t_n) \subset W_{M(t_n)}$. Thus M hypothesizes a proper superset of L on input t_n and hence M fails to learn L conservatively. This contradicts the choice of M, so φ_k is total.

Consequently, our procedure decides Tot, i. e., Tot is K-recursive. As this is impossible, we have $\mathcal{C}_{\mathrm{bc}} \notin ConsvTxt_{\mathrm{r.e.}}[K]$. □

Finally, thus Theorem 8 is proven, too. This is an example for the advantages of our characterisations; verifying Theorem 8 without Theorems 6 and 9 would have been possible, but more complicated. So the features of Gold-style learning can be exploited in the context of query learning.

Moreover, note that the indexable classes $\mathcal{C}_{\mathrm{lim}}$ and $\mathcal{C}_{\mathrm{bc}}$ defined in the proofs of Theorem 7 and Theorem 10, respectively, belong to $BcTxt_{\mathrm{r.e.}} \setminus LimTxt_{\mathrm{rec}}$. Up to now, the literature has not offered many such classes. The first example can be found in the seminal paper by Angluin [1], but its definition is quite involved and uses a diagonalisation. In contrast to that, $\mathcal{C}_{\mathrm{lim}}$ and $\mathcal{C}_{\mathrm{bc}}$ are defined compactly and explicitly without a diagonal construction and are—to the authors' knowledge—the first such classes known in $BcTxt_{\mathrm{r.e.}} \setminus LimTxt_{\mathrm{rec}}$.

6 Analogues in the world of learning with oracles

In our characterisations we have seen that the capability of query learners strongly depends on the hypothesis space and thus on the demands concerning the capabilities of the teacher. Of course it is more demanding to answer questions with respect to some uniformly r. e. family than to answer them with respect to some uniformly recursive family. In general, answering queries of the first kind might require the ability to solve the halting problem with respect to some Gödel numbering. In other words, the learner might use such queries to obtain access to an *oracle* for the halting problem.

In Theorem 9 we have seen that the idea of learners accessing oracles may be very useful in the context of our characterisations. In the sequel, we will show that this observation is not coincidental, i. e., we further analyse the relations of our previously considered inference types to inference types resulting from learning with the help of oracles.

The problem we consider in the following is to specify non-recursive sets $A \subseteq \mathbb{N}$ such that A-recursive query learners using uniformly recursive families as hypothesis spaces are as powerful as recursive learners using uniformly r. e. or uniformly K-r. e. families. For instance, we know that $rDisQ_{\text{rec}} \subset rDisQ_{\text{r.e.}} = LimTxt_{\text{rec}}$. So we would like to specify a set A, such that $LimTxt_{\text{rec}}$ equals the collection of all indexable classes which can be identified with A-recursive $rDisQ_{\text{rec}}$-learners. The latter collection will be denoted by $rDisQ_{\text{rec}}[A]$. Subsequently, similar notions are used correspondingly. Most of the claims below use K-recursive or Tot-recursive learners, where $K = \{i \mid \varphi_i(i) \text{ is defined}\}$ and $Tot = \{i \mid \varphi_i \text{ is a total function}\}$.

In the Gold-style model, the use of oracles has been analysed for example by Stephan [20], thus revealing a correspondence between conservative learning and learning in the limit:

Lemma 11 (Stephan [20]) $ConsvTxt_{\text{rec}}[K] = LimTxt_{\text{rec}}$.

Interestingly, Theorem 7, Theorem 9, and Lemma 11 imply $ConsvTxt_{\text{rec}}[K] \subset ConsvTxt_{\text{r.e.}}[K]$, although $ConsvTxt_{\text{rec}}$ and $ConsvTxt_{\text{r.e.}}$ are equal.

Relating learning in the limit to behaviourally correct learning, the use of oracles is illustrated by Lemma 12.

Lemma 12 *(1)* $ConsvTxt_{\text{rec}}[Tot] = LimTxt_{\text{rec}}[K] = BcTxt_{\text{r.e.}}$.
(2) $BcTxt_{\text{r.e.}}[A] = BcTxt_{\text{r.e.}}$ *for all* $A \subseteq \mathbb{N}$.

Proof. ad 2. Let $A \subseteq \mathbb{N}$. By definition $BcTxt_{\text{r.e.}} \subseteq BcTxt_{\text{r.e.}}[A]$. Thus it remains to prove the opposite inclusion, namely $BcTxt_{\text{r.e.}}[A] \subseteq BcTxt_{\text{r.e.}}$. For that purpose let $\mathcal{C} \in BcTxt_{\text{r.e.}}[A]$ be an indexable class. Fix an A-recursive IIM M such that $\mathcal{C}$ is $BcTxt_{\text{r.e.}}$-learned by M. Moreover, let $(L_i)_{i \in \mathbb{N}}$ be an indexing of $\mathcal{C}$.

Since M is a $BcTxt$-learner for each language L_i, there must be $BcTxt$-locking sequences σ_i for M, L_i, and some hypothesis space $\mathcal{H}$. Then M cannot learn any language L with $content(\sigma_i) \subseteq L \subset L_i$ for some i. Since M learns $\mathcal{C}$, this implies that no such L exists in $\mathcal{C}$ and thus, for each i, $content(\sigma_i)$ is a telltale for L_i. Hence $\mathcal{C}$ possesses a family of telltales and is $BcTxt_{\text{r.e.}}$-learnable. This yields $BcTxt_{\text{r.e.}}[A] = BcTxt_{\text{r.e.}}$.

ad 1. The proofs of $ConsvTxt_{\text{rec}}[Tot] \subseteq BcTxt_{\text{r.e.}}$, $LimTxt_{\text{rec}}[K] \subseteq BcTxt_{\text{r.e.}}$ are obtained by similar means as the proof of 2. It suffices to use Theorem 2 for $ConsvTxt_{\text{rec}}$ and $LimTxt_{\text{rec}}$ instead of the accordant statement for $BcTxt_{\text{r.e.}}$. Note that $LimTxt_{\text{rec}}[K] = BcTxt_{\text{r.e.}}$ has already been verified by Baliga et al. [4].

Next we prove $BcTxt_{\text{r.e.}} \subseteq ConsvTxt_{\text{rec}}[Tot]$ and $BcTxt_{\text{r.e.}} \subseteq LimTxt_{\text{rec}}[K]$.

For that purpose, let $\mathcal{C}$ be an indexable class in $\mathit{BcTxt}_{\text{r.e.}}$. By Theorem 2 there is an indexing $(L_i)_{i\in\mathbb{N}}$ of $\mathcal{C}$ which possesses a family of telltales. Next we show:

(i) $(L_i)_{i\in\mathbb{N}}$ possesses a *Tot*-recursively generable (uniformly K-r. e.) family of telltales.

(ii) A $\mathit{ConsvTxt}_{\text{rec}}$-learner ($\mathit{LimTxt}_{\text{rec}}$-learner) for $\mathcal{C}$ can be computed from any recursively generable (uniformly r. e.) family of telltales for $(L_i)_{i\in\mathbb{N}}$.

To prove (i), let for any $i \in \mathbb{N}$ a function f_i enumerate a set T_i as follows.

- $f_i(0) = \omega_z$ for $z = \min\{x \mid \omega_x \in L_i\}$.
- If $f_i(0), \ldots, f_i(n)$ are computed, then test whether or not there is some $j \in \mathbb{N}$ (some $j \leq n$), such that $\{f_i(0), \ldots, f_i(n)\} \subseteq L_j \subset L_i$.
  (* Note that this test is *Tot*-recursive (K-recursive). *)
- If such a number j exists, then $f_i(n+1) = \omega_z$ for $z = \min\{x \mid \omega_x \in L_i \setminus \{f_i(0), \ldots, f_i(n)\}\}$. If no such number j exists, then $f_i(n+1) = f_i(n)$.

With $T_i = \{f_i(x) \mid x \in \mathbb{N}\}$, it is not hard to verify that $(T_i)_{i\in\mathbb{N}}$ is a *Tot*-recursively generable (uniformly K-r. e.) family of telltales for $(L_i)_{i\in\mathbb{N}}$. Here note that, in the case of using a *Tot*-oracle, $T_i = \{f_i(x) \mid f_i(y{+}1) \neq f_i(y)$ for all $y < x\}$.

Finally, (ii) holds since Theorem 2.1/2.2 has a constructive proof, see Angluin [1] and Lange and Zeugmann [13].

Claims (i) and (ii) imply $\mathcal{C} \in \mathit{ConsvTxt}_{\text{rec}}[\mathit{Tot}]$ and $\mathcal{C} \in \mathit{LimTxt}_{\text{rec}}[K]$. So $\mathit{BcTxt}_{\text{r.e.}} \subseteq \mathit{ConsvTxt}_{\text{rec}}[\mathit{Tot}]$ and $\mathit{BcTxt}_{\text{r.e.}} \subseteq \mathit{LimTxt}_{\text{rec}}[K]$. □

Since the proofs of Lemma 11 and Lemma 12 are constructive as are the proofs of our characterisations above, we can deduce results like for example $\mathit{rDisQ}_{\text{rec}}[K] = \mathit{LimTxt}_{\text{rec}}$: Given $\mathcal{C} \in \mathit{LimTxt}_{\text{rec}}$, a K-recursive conservative IIM for $\mathcal{C}$ can be constructed from a $\mathit{LimTxt}_{\text{rec}}$-learner for $\mathcal{C}$. Moreover, an $\mathit{rDisQ}_{\text{rec}}$-learner for $\mathcal{C}$ can be constructed from a conservative IIM for $\mathcal{C}$. Now one can show that the latter relation can be lifted to the context of learning with K-recursive machines. That means, a K-recursive $\mathit{rDisQ}_{\text{rec}}$-learner for $\mathcal{C}$ can be constructed from a K-recursive conservative IIM for $\mathcal{C}$ and thus from a $\mathit{LimTxt}_{\text{rec}}$-learner. Similar results are obtained by combining Lemma 12 with our characterisations above. This proves the following theorem.

Theorem 13 *(1)* $\mathit{rSupQ}_{\text{rec}}[K] = \mathit{rDisQ}_{\text{rec}}[K] = \mathit{LimTxt}_{\text{rec}}$.
(2) $\mathit{rSupQ}_{\text{rec}}[\mathit{Tot}] = \mathit{rDisQ}_{\text{rec}}[\mathit{Tot}] = \mathit{rSupQ}_{\text{r.e.}}[\mathit{Tot}] = \mathit{rDisQ}_{\text{r.e.}}[\mathit{Tot}] = \mathit{BcTxt}_{\text{r.e.}}$.
(3) $\mathit{rSupQ}_{K\text{-r.e.}}[A] = \mathit{rDisQ}_{K\text{-r.e.}}[A] = \mathit{BcTxt}_{\text{r.e.}}$ *for all* $A \subseteq \mathbb{N}$.

But such arguments have to be used very carefully. Note that $\mathit{LimTxt}_{\text{rec}}[K] = \mathit{BcTxt}_{\text{r.e.}}$ and $\mathit{rDisQ}_{\text{r.e.}} = \mathit{LimTxt}_{\text{rec}}$, where both relations can be verified constructively. That means, given $\mathcal{C} \in BcTxt_{\text{r.e.}}$, a K-recursive $LimTxt_{\text{rec}}$-learner for $\mathcal{C}$ can be constructed from a $BcTxt_{\text{r.e.}}$-learner for $\mathcal{C}$. Moreover, an $rDisQ_{\text{r.e.}}$-learner for $\mathcal{C}$ can be constructed from a $LimTxt_{\text{rec}}$-learner for $\mathcal{C}$.

But still $BcTxt_{\text{r.e.}}$ is not equal to $rDisQ_{\text{r.e.}}[K]$. The reason is that in general a K-recursive $rDisQ_{\text{r.e.}}$-learner cannot be constructed from a K-recursive $LimTxt_{\text{rec}}$-learner, although the corresponding relation holds for recursive learners. In other words, the simulation of $LimTxt_{\text{rec}}$-learners using $rDisQ_{\text{r.e.}}$-learners cannot be lifted to the context of learning with K-recursive machines.

By the way, it is not hard to prove that $rDisQ_{\text{r.e.}}[K] = rDisQ_{\text{r.e.}}$: each access to a K-oracle can be simulated by a restricted disjointness query for some language W_i which is either empty or equal to Σ^*, similar to the method used in the proof of Theorem 9. Thus, when learning with restricted disjointness queries in uniformly r. e. families, query learners do not benefit from an additional access to an oracle for the halting problem. Obviously, the same holds for learning with restricted superset queries. Thus we obtain the following theorem.

Theorem 14 *(1)* $rSupQ_{\text{r.e.}}[K] = rSupQ_{\text{r.e.}}$.
(2) $rDisQ_{\text{r.e.}}[K] = rDisQ_{\text{r.e.}}$.

7 Conclusions

We have considered prototypical formal models of machine intelligence using ideas of algorithmic learning theory. The essential aspect of our analysis has been the approach of incorporating different types of scenarios based on different types of learning processes. The formalisation of the corresponding models has helped to describe important parameters of such scenarios, the modification of which can affect the capabilities of the corresponding learning algorithms.

One such parameter is for instance the hypothesis space used, which may be seen as a kind of representation scheme a teacher and a learner use for communication. [2] Its relevance has been approved theoretically, e. g., by separations of the query learning models resulting from different types of hypothesis spaces.

Particularly in the query learning model, the hypothesis space influences the requirements on the teacher and thus the amount of information the learner is presented, as well.

[2] Note that even Gold's model can be interpreted in terms of an interaction between a teacher and a learner. Here the teacher is simply a device presenting the examples to the learner.

The amount and form of information as well as the procedural constraints (limiting versus one-shot learning processes) constitute other relevant parameters, as expressed in the trade-off between two formal approaches to language learning.

In this context we have focussed on a comparison of *Gold-style language learning* (see Gold [9])—interpreting learning as a *limiting process* in which the learner may change its mind arbitrarily often before converging to a correct hypothesis—to *language learning via queries* (see Angluin [2,3])—interpreting learning as a *one-shot process* in which the learner is required to identify the target concept with just one hypothesis.

Although these two approaches seem rather unrelated at first glance, first results in [14,15] have shown that there are still some common features concerning the structure of learnable classes and the algorithmic complexity of learners. Combining Gold-style learners with query learners, Jain and Kinber [11,12] have moreover addressed the question of how to exploit the capabilities of either type of learners in the context of the other type.

Following the line of observations in [14,15], we have now provided characterisations of different models of Gold-style learning (learning in the limit, conservative inference, and behaviourally correct learning) in terms of query inference. Thus we have described the circumstances which are necessary to replace limit learners by at least equally powerful one-shot learners. In order to do so, the crucial parameters are the type of queries (restricted superset or restricted disjointness queries) and the underlying hypothesis space (uniformly recursive, uniformly r. e., or uniformly K-r. e. families). The characterisations of Gold-style language learning have been formulated in dependence of these parameters; the results have been presented in Section 4.

This analysis has lead to an important observation, namely that there is a natural query learning type hierarchically in-between Gold-style learning in the limit and behaviourally correct learning. Astonishingly, this query learning type could then again be characterised in terms of Gold-style inference, as has been shown in Section 5.

The latter characterisation has revealed another important relation of the inference types we have considered, namely an interconnection with non-recursive learners (learners using oracles). That this interconnection is not at all coincidental, has been demonstrated in Section 6.

As a consequence of our characterisations, the knowledge about the learnability of particular classes in either model may be used immediately to bring forward statements in the other related model. For instance, since a very prominent indexable class, namely the class of all erasing pattern languages, is not $BcTxt_{\text{r.e.}}$-learnable (see Reidenbach [18] for the corresponding definitions and results), we can conclude that no kind of oracle may help to learn this class with superset queries, even if uniformly K-r. e. hypothesis spaces are used for communication.

Acknowledgements

We would like to express our gratitude to the anonymous referees, whose detailed and critical reviews on former versions of this paper account for many improvements of the overall presentation.

References

[1] Angluin, D. (1980), Inductive inference of formal languages from positive data, *Information and Control* 45, 117–135.

[2] Angluin, D. (1988), Queries and concept learning, *Machine Learning* 2, 319–342.

[3] Angluin, D. (2004), Queries revisited, *Theoretical Computer Science* 313, 175–194.

[4] Baliga, G., Case, J., and Jain, S. (1999), The synthesis of language learners, *Information and Computation* 152, 16–43.

[5] Blum, M. (1967), A machine-independent theory of the complexity of recursive functions, *Journal of the ACM* 14, 322–336.

[6] Blum, L., and Blum, M. (1975), Toward a mathematical theory of inductive inference, *Information and Control* 28, 125–155.

[7] Case, J., and Lynes, C. (1982), Machine inductive inference and language identification, In: *Proc. 9th International Colloquium on Automata, Languages and Programming 1982*, Lecture Notes in Computer Science 140, pp. 107–115, Springer-Verlag.

[8] Fulk, M. (1990), Prudence and other conditions on formal language learning, *Information and Computation* 85, 1–11.

[9] Gold, E. M. (1967), Language identification in the limit, *Information and Control* 10, 447–474.

[10] Hopcroft, J. E., and Ullman, J. D. (1979), *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley Publishing Company.

[11] Jain, S., and Kinber, E. (2004), Learning languages from positive data and negative counterexamples, In: *Proc. 15th International Conference on Algorithmic Learning Theory*, Lecture Notes in Artificial Intelligence 3244, pp. 54–68, Springer-Verlag.

[12] Jain, S., and Kinber, E. (2004), Learning languages from positive data and a finite number of queries, Technical Report TRC4/04, National University of Singapore, (a preliminary version has appeared in Proc. 24th International Conference on Foundations of Software Technology and Theoretical Computer Science).

[13] Lange, S., and Zeugmann, T. (1993), Language learning in dependence on the space of hypotheses, In: *Proc. 6th Annual Conference on Computational Learning Theory*, pp. 127–136, ACM Press.

[14] Lange, S., and Zilles, S. (2003), On the learnability of erasing pattern languages in the query model, In: *Proc. 14th International Conference on Algorithmic Learning Theory*, Lecture Notes in Artificial Intelligence 2842, pp. 129–143, Springer-Verlag.

[15] Lange, S., and Zilles, S. (2004), Formal language identification: query learning vs. Gold-style learning, *Information Processing Letters* 91, 285–292.

[16] Lange, S., and Zilles, S. (2004), Replacing limit learners with equally powerful one-shot query learners, In: *Proc. 17th Annual Conference on Learning Theory*, Lecture Notes in Artificial Intelligence 3120, pp. 155–169, Springer-Verlag.

[17] Lange, S., and Zilles, S. (2004), Comparison of query learning and Gold-style learning in dependence of the hypothesis space, In: *Proc. 15th International Conference on Algorithmic Learning Theory*, Lecture Notes in Artificial Intelligence 3244, pp. 99–113, Springer-Verlag.

[18] Reidenbach, D. (2002), A negative result on inductive inference of extended pattern languages, In: *Proc. 13th International Conference on Algorithmic Learning Theory*, Lecture Notes in Artificial Intelligence 2533, pp. 308–320, Springer-Verlag.

[19] Rogers, H. Jr. (1987), *Theory of Recursive Functions and Effective Computability*, MIT Press.

[20] Stephan, F. (1999), *Degrees of Computing and Learning*, Habilitationsschrift, Ruprecht-Karls-Universität, Heidelberg.

[21] Zeugmann, T., and Lange, S. (1995), A guided tour across the boundaries of learning recursive languages, In: *Algorithmic Learning for Knowledge-Based Systems*, Lecture Notes in Artificial Intelligence 961, pp. 190–258, Springer-Verlag.