On Batch Teaching Without Collusion*

Shaun Fallat   SHAUN.FALLAT@UREGINA.CA
Department of Mathematics and Statistics, University of Regina

David Kirkpatrick   KIRK@CS.UBC.CA
Department of Computer Science, University of British Columbia

Hans U. Simon   HSIMON@MPI-INF.MPG.DE
Max Planck Institute for Informatics

Abolghasem Soltani   ASOLTANI@UREGINA.CA
Department of Computer Science, University of Regina

Sandra Zilles   ZILLES@CS.UREGINA.CA
Department of Computer Science, University of Regina

Editor: Daniel Hsu

Abstract

Formal models of learning from teachers need to respect certain criteria to avoid collusion. The most commonly accepted notion of collusion-avoidance was proposed by Goldman and Mathias (1996), and various teaching models obeying their criterion have been studied. For each model M and each concept class $\mathcal{C}$, a parameter M-$\text{TD}(\mathcal{C})$ refers to the *teaching dimension* of concept class $\mathcal{C}$ in model M—defined to be the number of examples required for teaching a concept, in the worst case over all concepts in $\mathcal{C}$.

This paper introduces a new model of teaching, called no-clash teaching, together with the corresponding parameter $\text{NCTD}(\mathcal{C})$. No-clash teaching is provably optimal in the strong sense that, given *any* concept class $\mathcal{C}$ and *any* model M obeying Goldman and Mathias's collusion-avoidance criterion, one obtains $\text{NCTD}(\mathcal{C}) \leq M$-$\text{TD}(\mathcal{C})$. We also study a corresponding notion NCTD^+ for the case of learning from positive data only, establish useful bounds on NCTD and NCTD^+, and discuss relations of these parameters to other complexity parameters of interest in computational learning theory.

We further argue that Goldman and Mathias's collusion-avoidance criterion may in some settings be too weak in that it admits certain forms of interaction between teacher and learner that could be considered collusion in practice. Therefore, we introduce a strictly stronger notion of collusion-avoidance and demonstrate that the well-studied notion of Preference-based Teaching is optimal among all teaching schemes that are strongly collusion-avoiding on all finite subsets of a given concept class.

Keywords: machine teaching, collusion-freeness, VC dimension, teaching dimension, sample compression

---

*. Some of the material in this paper appeared in preliminary form in (Kirkpatrick et al., 2019)

1. Introduction

Models of machine learning from carefully chosen examples, i.e., from teachers, have gained increased interest in recent years, due to various application areas, such as robotics (Argall et al., 2009), trustworthy AI (Zhu et al., 2018), and pedagogy (Shafto et al., 2014). Machine teaching is also related to inverse reinforcement learning (Ho et al., 2016), to sample compression (Moran et al., 2015; Doliwa et al., 2014), and to curriculum learning (Bengio et al., 2009).

The paper at hand is concerned with abstract notions of teaching, as studied in computational learning theory. In particular, it assumes that the learning problem is given by a fixed concept class $\mathcal{C}$, of which an unknown target concept C is chosen and has to be identified by a learner from partial information about C. A teacher and a learner are viewed as two mappings defined on the basis of the given concept class $\mathcal{C}$. The teacher compresses each concept C in $\mathcal{C}$ into a set or sequence of correctly labelled examples, i.e., of pairs $(x, C(x))$, where x is an element of the underlying domain and $C(x)$ is a binary label indicating whether or not x belongs to C. The learner in return is presented with a set or sequence of such labelled examples and outputs a concept in $\mathcal{C}$. By contrast with models of learning from randomly chosen examples, the examples presented to the learner are more helpful, as the teacher is assumed to be *benevolent*.

A variety of formal models of teaching have been proposed in the literature, for example, the classical teaching dimension model (Goldman and Kearns, 1995), the optimal teacher model (Balbach, 2008), recursive teaching (Zilles et al., 2011), or preference-based teaching (Gao et al., 2017). In each of these models, the teacher mapping T assigns a finite set $T(C)$ of correctly labelled examples to a concept C in a concept class $\mathcal{C}$ in a way that the learner can reconstruct C from $T(C)$. In particular, they are models of *batch teaching*, i.e., of teaching and learning using *sets* of labeled examples, and stand in contrast with models of *sequential* teaching (Mansouri et al., 2019), in which concepts are encoded in *sequences* of examples.

An interesting variant of all of these models is obtained when disallowing negative examples in teaching. Learning from positive examples only has been studied extensively in the computational learning theory literature, see, e.g., (Denis, 2001; Angluin, 1980) and is motivated by studies on language acquisition (Wexler and Culicover, 1980) or, more recently, by problems of learning user preferences from a user's interactions with, say, an e-commerce system (Schwab et al., 2000), as well as by problems in bioinformatics (Wang et al., 2006).

Intuitively, unfair collusion between the teacher and the learner should not be allowed in any formal model of teaching. For example, one would not want the teacher and learner to agree on a total order over the domain and a total order over the concept class and then to simply use the ith instance in the domain for teaching the ith concept, irrespective of the actual structure of the concept class. However, there is no general definition of what constitutes collusion, and of what constitutes desirable or undesirable forms of learning. In this manuscript, we focus on two notions of collusion-avoidance. The first, which we refer to as *weak* collusion-avoidance, is the property proposed by Goldman and Mathias (1996) that has served as a foundation for a majority of the teaching models studied in the literature. The second, more restrictive, notion we refer to as *strong* collusion-avoidance. In a nutshell, weak collusion-avoidance demands of a teacher mapping T that it admits a learner L that is *persistent* in the sense that $L(S) = C$ for all sets S of labelled examples that include $T(C)$ and are consistent with C. (In other words, adding more information about C to $T(C)$ will not divert the learner to a different hypothesis.) Strong collusion-avoidance on a set

$\mathcal{C}$ of concepts demands that, in addition, the teacher mapping T admits a unique injective learner mapping L on $T(\mathcal{C})$ such that $L(T(\mathcal{C})) \subseteq \mathcal{C}$.

Most existing abstract models of machine teaching are known to satisfy the Goldman-Mathias (weak) collusion-avoidance criterion. Historically, some of these models were designed in order to overcome weaknesses of the previous models. For example, the optimal teacher model by Balbach (2008) is designed to overcome limitations of the classical teaching dimension model, and was likewise superseded by the recursive teaching model (Zilles et al., 2011). The latter again failed to capture teachability of many interesting infinite concept classes, which gave rise to the model of preference-based teaching (Gao et al., 2017). Each model strictly dominates the previous one in terms of the *teaching complexity*, i.e., the worst-case number of examples needed for teaching a concept in the underlying concept class $\mathcal{C}$. In this context, one quite natural question has been ignored in the literature to date: what is the smallest teaching complexity that can be achieved while respecting the Goldman-Mathias's collusion-avoidance criterion? This is one of the central questions addressed in this paper.

1.1 Overview of Results

Our first contribution is the formal definition of a teaching model that has, for every concept class $\mathcal{C}$, the provably smallest teaching complexity among all teaching models that satisfy the Goldman-Mathias collusion-avoidance criterion. We call this model *no-clash teaching*, since its core property, which turns out to be characteristic for the Goldman-Mathias collusion-avoidance criterion, requires that no pair of concepts are consistent with the union of their teaching sets. A similar property was used once in the literature in the context of sample compression schemes (Kuzmin and Warmuth, 2007), and dubbed the *non-clashing* property.

For example, consider a concept class (i.e., set system) $\mathcal{C}_A$ over the instance space $\{0, 1, 2, 3\}$, consisting of the four concepts of the form $\{i, (i + 1) \bmod 4\}$ for $0 \leq i \leq 3$. Then no-clash teaching is possible by assigning the singleton set $\{(i, 1)\}$ (interpreted as the information "i belongs to the target concept") as a teaching set to the concept $\{i, (i + 1) \bmod 4\}$; no two distinct concepts are consistent with the union of their assigned teaching sets. Thus, in the no-clash setting, each concept in $\mathcal{C}_A$ can be taught with a single example. By comparison, consider the classical teaching dimension model, in which a teaching set for a given concept is required to be inconsistent with all other concepts in the concept class (Goldman and Kearns, 1995). It is not hard to see that, under such constraints, no concept in $\mathcal{C}_A$ can be taught with a single example; a smallest teaching set for concept $\{i, (i + 1) \bmod 4\}$ would then be $\{(i, 1), ((i + 1) \bmod 4, 1)\}$.

We call the worst-case number of examples needed for non-clashing teaching of any concept C in a given concept class $\mathcal{C}$ the *no-clash teaching dimension* of $\mathcal{C}$, abbreviated $\mathrm{NCTD}(\mathcal{C})$, and we study a variant $\mathrm{NCTD}^+(\mathcal{C})$ in which teaching uses only positive examples. In the example above, $\mathrm{NCTD}(\mathcal{C}_A) = \mathrm{NCTD}^+(\mathcal{C}_A) = 1$, while the classical teaching dimension is 2.

As additional examples consider the infinite concept classes $\mathcal{C}_B$ (respectively, $\mathcal{C}_{B+}$, $\mathcal{C}_{B-}$) over the real numbers consisting of concepts which are half-open (respectively, closed, fully open) intervals of the form $[x, x + 1)$ (respectively, $[x, x + 1]$, $(x, x + 1)$), where $x \in \mathbb{R}$. In the classical teaching dimension model no concept in $\mathcal{C}_B$ can be taught with finitely many examples. However, both $\mathcal{C}_{B+}$ and $\mathcal{C}_{B-}$ have finite classical teaching dimension (two and three, respectively). In the no-clash setting, concepts in all three classes can be taught with just a single example.

The value $\mathrm{NCTD}(\mathcal{C})$ being the smallest teaching complexity parameter of $\mathcal{C}$ satisfying the Goldman-Mathias collusion-avoidance criterion makes it interesting for several reasons.

(1) NCTD represents the limit of data efficiency in teaching when respecting Goldman and Mathias's weak notion of collusion-avoidance. Therefore the study of NCTD has the potential to further our understanding how collusion-avoidance constrains teaching. It will also serve to compare other notions of collusion-avoidance (see, e.g., (Zilles et al., 2011)) to that of Goldman and Mathias.

(2) An open question in computational learning theory is whether the VC-dimension (VCD), (Vapnik and Chervonenkis, 1971), which characterizes the sample complexity of learning from randomly chosen examples, also characterizes teaching complexity for some reasonable notion of batch teaching. Over the past decade, the first strong connections between batch teaching and VCD were established, culminating in an upper bound on the recursive teaching dimension (RTD) that is quadratic in VCD (Hu et al., 2017), but it remains open whether this bound can be improved to be linear in VCD. Obviously, now NCTD is a much stronger candidate for a linear relationship with VCD than RTD is. In fact, there is no concept class known yet for which NCTD exceeds VCD. The only known teaching complexity notion that is bounded from above by VCD (and also the only one known to be bounded from above by a function linear in VCD) stems from a model of sequential teaching and was introduced by Mansouri et al. (2019). Thus, NCTD is to date the only known candidate for a batch teaching parameter that has an upper bound in $O(\mathrm{VCD})$.

(3) The problem of relating teaching complexity to VCD is connected to the long-standing open problem of determining whether VCD is an upper bound on the size of the smallest possible sample compression scheme (Littlestone and Warmuth, 1986; Floyd and Warmuth, 1995) of a concept class. Some interesting relations between sample compression and teaching have been established for RTD (Moran et al., 2015; Doliwa et al., 2014; Darnstädt et al., 2016). The study of NCTD can potentially strengthen such relations.

To sum up, our new notion of non-clashing teaching, while respecting the Goldman-Mathias collusion-avoidance criterion, is of relevance to the study of important problems in computational learning theory.

Interestingly, non-clashing teaching also reveals a weakness of Goldman and Mathias's notion of collusion-avoidance. Consider for example the concept class $\mathcal{C}_A$ from above, i.e., the class consisting of the concepts $\{i, (i+1) \bmod 4\}$, $0 \leq i \leq 3$, over the instance space $\{0, 1, 2, 3\}$. As stated above, an optimal non-clashing teacher is obtained when assigning the singleton set $\{(i, 1)\}$ to the concept $\{i, (i+1) \bmod 4\}$. Now it is not hard to see that the mapping that assigns $\{((i+1) \bmod 4, 1)\}$ to the concept $\{i, (i+1) \bmod 4\}$ is also an optimal non-clashing teacher for $\mathcal{C}_A$. A learner will first have to disambiguate between various potential teachers in order to be able to successfully decode a given teaching set. Arguably, the information exchange required for this disambiguation could be considered some form of collusion—in particular, a form of collusion not forbidden by Goldman-Mathias collusion-avoidance.

This is where the stronger notion of collusion-avoidance comes into play. For a teacher T over finite concept class $\mathcal{C}$, it stipulates that no alternate teacher for $\mathcal{C}$ has the same range as T.[1] A major result is that the preference-based teaching dimension (PBTD, (Gao et al., 2017)), which is a well-known complexity parameter of batch teaching and coincides with RTD on finite concept classes, is exactly the complexity of teacher mappings that are strongly collusion-avoiding on all finite subsets of the given concept class.

---

1. For infinite concept classes, the definition is just slightly more involved—see Section 5.

Thus, two major contributions of our paper are (i) to characterize the optimal complexity of weakly collusion-avoiding teaching by introducing non-clashing teaching, and (ii) to address weaknesses of this notion by introducing strong collusion-avoidance, which in turn serves as a characterization of preference-based teaching (and, in the finite case, of recursive teaching).

1.2 A Guide to What Follows

Section 2 provides background on formal models of teaching that satisfy the weak (Goldman-Mathias) collusion-avoidance criterion, together with their associated teaching complexity (teaching dimension) parameters. This motivates the question: what is the most permissive (and hence lowest complexity) teaching model that satisfies the weak collusion avoidance constraints.

This is answered in Section 3 with the introduction of non-clashing teaching and its associated teaching complexity parameter NCTD. Some the basic properties of this complexity parameter, including lower bounds, are developed within Subsection 3.2. Section 4 explores the relationship between non-clashing teaching and other learning-theoretic notions, including VC-dimension and sample compression.

Section 5 introduces strong collusion avoidance and establishes its close relationship with preference-based teaching.

Finally, Section 6 revisits all of the formal teaching notions that we have considered, including non-clashing teaching, under the unifying frameworks of constrained preferences, constrained bipartite matchings, and broadened recursion.

2. Preliminaries

Given a domain $\mathcal{X}$, a concept over $\mathcal{X}$ is a subset $C \subseteq \mathcal{X}$. We usually denote by $\mathcal{C}$ a *concept class* over $\mathcal{X}$, i.e., a set of concepts over $\mathcal{X}$. Implicitly, we identify a concept C over $\mathcal{X}$ with a mapping $C : \mathcal{X} \to \{0, 1\}$, where $C(x) = 1$ iff $x \in C$.

Definition 1 *A labelled example is a pair $(x, \ell) \in \mathcal{X} \times \{0, 1\}$. An example with the label $\ell = 1$ is a positive example, while $\ell = 0$ is the label of a negative example. Let S be any finite set of labelled examples over $\mathcal{X}$, and $\mathcal{C}$ any concept class over $\mathcal{X}$. Then S and concept $C \in \mathcal{C}$ are said to be (mutually) consistent, if for every pair $(x, \ell) \in S$, $\ell = C(x)$. If S is consistent with at least one concept in $\mathcal{C}$ then S is called a $\mathcal{C}$-sample.*

Intuitively, the notion of teaching refers to compressing any concept C in a given concept class $\mathcal{C}$ to a $\mathcal{C}$-sample consistent with C, by means of an invertible mapping.

Definition 2 (Teacher Mapping) *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$. A teacher mapping for $\mathcal{C}$ is an injective mapping T from $\mathcal{C}$ to the set of $\mathcal{C}$-samples such that, for all $C \in \mathcal{C}$, $T(C)$ is consistent with C. The order of T on $\mathcal{C}$, denoted by $\mathrm{ord}(T, \mathcal{C})$, is then defined by $\mathrm{ord}(T, \mathcal{C}) = \max\{|T(C)| \mid C \in \mathcal{C}\}$. A teacher mapping T is called positive on $\mathcal{C}$ if $T(C) \subseteq \mathcal{X} \times \{1\}$ for all $C \in \mathcal{C}$.*

Definition 3 (Learner Mapping) *A learner mapping for $\mathcal{C}$ is a partial function L on the set of $\mathcal{C}$-samples to concepts in $\mathcal{C}$ such that if L is defined on S then S is consistent with $L(S)$. A learner mapping L is persistent on $\mathcal{C}$ if, for any $\mathcal{C}$-sample S in the domain of L and any superset S' of*

S that is consistent with $L(S)$, $L(S') = L(S)$.[2] *A learner mapping L for $\mathcal{C}$, forms a* successful teacher-learner pair *with the teacher mapping T on the subset $\mathcal{C}' \subseteq \mathcal{C}$ if $L(T(C)) = C$, for all $C \in \mathcal{C}'$.*

While there is no objective measure of how "reasonable" a formal model of teaching is, the literature offers some notions of what constitutes an "acceptable" model of teaching, i.e., one in which the teacher and learner do not collude. So far, the property of collusion-avoidance that found the most positive resonance in the literature is the one defined by Goldman and Mathias. Intuitively, Goldman and Mathias's criterion captures the idea that a learner conjecturing concept C will not change its mind when given additional information consistent with C.

Definition 4 (Weak Collusion-Avoidance (Goldman and Mathias, 1996)) *A teacher mapping T for a concept class $\mathcal{C}$ is said to be* weakly collusion-avoiding *on $\mathcal{C}$ if there exists a persistent learner mapping L that forms a successful teacher-learner pair with T on $\mathcal{C}$.*

Remark 5 This definition deliberately does not require the persistent learner mapping L to be defined on all $\mathcal{C}$-samples. For illustration, consider the concept class $\mathcal{P}_2$, the power set over a domain $\{x_1, x_2\}$ of size 2. This class has a teacher mapping T satisfying the weak collusion-avoidance property, with $\text{ord}(T, \mathcal{P}_2) = 1$; for example, T can map the empty concept to $\{(x_1, 0)\}$, the concept $\{x_1\}$ to $\{(x_2, 0)\}$, the concept $\{x_2\}$ to $\{(x_2, 1)\}$, and finally the concept $\{x_1, x_2\}$ to $\{(x_2, 1)\}$. A persistent learner can form a successful teacher-learner pair with this teacher if it is not required to produce an output on the empty sample.

By contrast, weakly collusion-avoiding teaching of $\mathcal{P}_2$ is no longer possible with an order of 1 if, in addition, the persistent mapping L is required to be defined on all $\mathcal{C}$-samples. To see this, suppose, without loss of generality, that a persistent learner L for $\mathcal{P}_2$ returns the empty concept on input of the empty sample. When fed $\{(x_1, 0)\}$ or $\{(x_2, 0)\}$, this learner would have to output the empty concept as well, to be successful with a weakly collusion-avoiding teacher. Hence, L would need to see at least the example $(x_1, 1)$ to output the concept $\{x_1\}$, and it would need to see at least $\{(x_2, 1)\}$ to output the concept $\{x_2\}$. Thus, at least one of the three concepts $\{x_1\}$, $\{x_2\}$, $\{x_1, x_2\}$ would need to be taught using two examples, for L to allow for weakly collusion-avoiding teaching.

In the following we describe a natural *normal form* for teacher mappings that satisfy the weakly collusion-avoiding property. A teacher mapping T' is said to be an *extension* of T if $T(C) \subseteq T'(C)$ holds for every $C \in \mathcal{C}$. Clearly, if T' is an extension of T and T is weakly collusion-avoiding then so to is T'. It follows immediately that:

Proposition 6 *If teacher mapping T is weakly collusion-avoiding then there is a weakly collusion-avoiding teacher mapping T' for which $|T'(C)| = \text{ord}(T, \mathcal{C})$, for all $C \in \mathcal{C}$.*

We will find it helpful to exploit this normal form as follows: if $B_{\mathcal{C}, \mathcal{X}}^{(k)}$ denotes the bipartite graph, with vertices associated with elements of $\mathcal{C}$ (black vertices) and $\mathcal{C}$-samples of size k over $\mathcal{X}$ (white vertices), and an edge from $C_i \in \mathcal{C}$ to $\mathcal{C}$-sample S_j whenever C_i is consistent with S_j, then any weakly collusion-avoiding teacher mapping of order k can be expressed as a matching in the bipartite graph $B_{\mathcal{C}, \mathcal{X}}^{(k)}$ that saturates all of the black vertices.

---

2. In the context of recursion-theoretic learning theory (Lange et al., 2008), persistent learner mappings are sometimes called *conservative*.

2.1 Definitive Teaching

The first model of teaching that was proposed in the literature required from a teacher mapping T that the concept $C \in \mathcal{C}$ be the only concept in $\mathcal{C}$ that is consistent with $T(C)$, for any $C \in \mathcal{C}$ (Shinohara and Miyano, 1991; Goldman and Kearns, 1995). This led to the definition of the well-known teaching dimension[3] parameter. The following presents this definition in a modified, but equivalent, form:

Definition 7 (DTD (Shinohara and Miyano, 1991; Goldman and Kearns, 1995)) *Fix a concept class $\mathcal{C}$ over a domain $\mathcal{X}$, and let $C \in \mathcal{C}$ be a concept. A* definitive teaching set *for C (with respect to the concept class $\mathcal{C}$) is a $\mathcal{C}$-sample S for which C is the only consistent concept in $\mathcal{C}$. By $\mathrm{DT}^{(k)}(\mathcal{C})$ we denote the subset of $\mathcal{C}$ consisting of concepts with definitive teaching sets (with respect to $\mathcal{C}$) of size at most k. The* definitive teaching dimension *of $\mathcal{C}$, denoted $\mathrm{DTD}(\mathcal{C})$, is then defined as $\min\{k \mid \mathrm{DT}^{(k)}(\mathcal{C}) = \mathcal{C}\}$ (with $\mathrm{DTD}(\mathcal{C}) = \infty$ if, for every k, there is some concept whose smallest definitive teaching set S has size greater than k). Further, we define*

$$\mathrm{DTD}_{\min}(\mathcal{C}) = \min\{k \mid \mathrm{DT}^{(k)}(\mathcal{C}) \neq \emptyset\}\,.$$

The positive definitive teaching dimension *of $\mathcal{C}$, denoted by $\mathrm{DTD}^{+}(\mathcal{C})$, is defined analogously to $\mathrm{DTD}(\mathcal{C})$, where the sets S are required to contain only positive examples.*

As we have already seen, there is no relationship between the size of a concept class and its definitive teaching dimension: clearly there exist uncountably large concept classes $\mathcal{C}$, each with uncountably many elements, with $\mathrm{DTD}(\mathcal{C}) = 1$. It is not hard to construct concept classes for which DTD^{+} is an arbitrarily large multiple of DTD, for example the set of concepts that include all but one instance from the underlying domain.

For example, let $\mathcal{C}_D$ be a concept class over a domain $\mathcal{X}$ of exactly m elements, containing the empty concept, all singleton concepts over $\mathcal{X}$, and no other concepts (Goldman and Kearns, 1995). Then $\mathrm{DT}^{(1)}(\mathcal{C}_D)$ consists of all of the singleton concepts, since $\{(x, 1)\}$ serves as a teaching set for $\{x\}$. However, $\mathrm{DTD}(\mathcal{C}_D) = m$, since any set of up to $m - 1$ negative examples is consistent with some singleton concept and hence any teaching set for the empty concept must include all m negative examples. Similarly, the concept class $\mathcal{C}_E$ over $\mathbb{N}$, containing the empty concept and all singleton concepts, but no others, has $\mathrm{DTD}(\mathcal{C}_E) = \infty$.

Note that implicit in the definition above is the existence of a weakly collusion-avoiding teacher mapping T of order at most $\mathrm{DTD}(\mathcal{C})$.

2.2 Recursive Teaching

As mentioned in the introduction, several relaxations of the notion of definitive teaching have been proposed in the literature.

The RTD, short for "recursive teaching dimension", is a well-studied teaching parameter defined by Zilles et al. (2011). It is determined by removing from the concept class $\mathcal{C}$ all those concepts that admit definitive teaching sets of size $\mathrm{DTD}_{\min}$ and then recursing on the remaining concepts. The largest $\mathrm{DTD}_{\min}$ value thus encountered is the RTD. The following provides an equivalent definition to that of Zilles et al. (2011):

---

3. In the following, in order to avoid confusion with other notions of teaching dimension, we have chosen to refer to this notion as the *definitive teaching dimension*.

Definition 8 (RTD) *Let $\mathcal{C}$ be a concept class. The largest subset $\mathrm{RT}_r^{(k)}(\mathcal{C})$ of $\mathcal{C}$ for which a recursive teacher mapping, using $\mathcal{C}$-samples of size at most k, can be formulated with at most r recursive steps, is defined as follows: $\mathrm{RT}_0^{(k)}(\mathcal{C}) = \mathrm{DT}^{(k)}(\mathcal{C})$ and for all $i \in \mathbb{N}$, $\mathrm{RT}_{i+1}^{(k)}(\mathcal{C}) = \mathrm{RT}_i^{(k)}(\mathcal{C}) \cup \mathrm{DT}^{(k)}(\mathcal{C} \setminus \mathrm{RT}_i^{(k)}(\mathcal{C}))$.*

Let $\mathrm{RT}_*^{(k)}(\mathcal{C}) = \cup_{r \geq 0} \mathrm{RT}_r^{(k)}(\mathcal{C})$. The recursive teaching dimension *of $\mathcal{C}$, denoted $\mathrm{RTD}(\mathcal{C})$ is defined as $\min\{k \mid \mathrm{RT}_*^{(k)}(\mathcal{C}) = \mathcal{C}\}$.*

The positive recursive teaching dimension *of $\mathcal{C}$, denoted by $\mathrm{RTD}^+(\mathcal{C})$, is defined analogously, when the $\mathcal{C}$-samples are required to contain only positive examples.*

Note that implicit in the definition of $\mathrm{RT}_*^{(k)}(\mathcal{C})$ is the existence of a weakly collusion-avoiding teacher mapping T, called a *minimum depth recursive teaching plan* for $\mathrm{RT}_*^{(k)}(\mathcal{C})$ of order k, that assigns to each element of $C \in \mathrm{DT}^{(k)}(\mathcal{C} \setminus \mathrm{RT}_i^{(k)}(\mathcal{C}))$ a definitive teaching set for C with respect to $\mathcal{C} \setminus \mathrm{RT}_i^{(k)}(\mathcal{C})$, of size at most k.

As with DTD, it is not hard to construct concept classes for which RTD^+ is an arbitrarily large multiple of RTD. Further, it should be clear from the definition that, for all k, $\mathrm{DT}^{(k)}(\mathcal{C}) = \mathrm{RT}_0^{(k)}(\mathcal{C}) \subseteq \mathrm{RT}_*^{(k)}(\mathcal{C})$, and hence $\mathrm{RTD}(\mathcal{C}) \leq \mathrm{DTD}(\mathcal{C})$. In fact, $\mathrm{RTD}^+(\mathcal{C}_D) = \mathrm{RTD}^+(\mathcal{C}_E) = 1$, illustrating the fact that, for some concept classes, RTD^+ can be an arbitrarily small fraction of DTD. If we define the concept classes $\mathcal{C}_F = \{\{i, i+1\} \mid i \in \mathbb{N}\}$ and $\mathcal{C}_G = \{\{i, i+1\} \mid i \in \mathbb{Z}\}$, as well as $\mathcal{C}_{F+} = \mathcal{C}_F \cup \{\emptyset\}$ and $\mathcal{C}_{G+} = \mathcal{C}_G \cup \{\emptyset\}$, it is straightforward to confirm that $\mathrm{RTD}(\mathcal{C}_F) = 1$, $\mathrm{RTD}(\mathcal{C}_G) = 2$, and $\mathrm{RTD}(\mathcal{C}_{F+}) = \mathrm{RTD}(\mathcal{C}_{G+}) = \infty$. Note, however, that if we generalize the notion of recursive teaching to 2-*recursive teaching* in which maximal size recursive teaching sets are taught in successive rounds, then the recursive teaching dimension of the concept class $\mathcal{C}_{F+}$ is reduced to 1.

2.3 Preference-Based Teaching

In the model of preference-based teaching, a preference relation $\succ$ on $\mathcal{C}$, known to both the teacher and the learner, is used to reduce the size of teaching sets. As with the recursive teaching model, a concept C need not be be the only concept consistent with its teaching set $T(C)$; to resolve ambiguity, C is chosen by the learner as the unique most preferred concept in $\mathcal{C}$ that is consistent with $T(C)$. While it is natural to express preferences among concepts using a strict order on $\mathcal{C}$, in order to ensure that the teacher mapping T is injective, it suffices that the preference relation $\succ$ is asymmetric on $\mathcal{C}$.

Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$ and $\succ$ any asymmetric binary relation on $\mathcal{C}$. We say that concept C is preferred over concept C' (with respect to $\succ$), if $C \succ C'$. A $\succ$-*distinguished* teaching set for C over $\mathcal{C}$ is a $\mathcal{C}$-sample S such that

1. S is consistent with C, and

2. $C \succ C'$ for all $C' \in \mathcal{C} \setminus \{C\}$ such that S is consistent with C'.

We define the set $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ)$ to be the set of all concepts $C \in \mathcal{C}$ that have a $\succ$-distinguished teaching set of size at most k over $\mathcal{C}$. The *preference-based teaching dimension* of $\mathcal{C}$ with respect to $\succ$, denoted by $\mathrm{PBTD}(\mathcal{C}, \succ)$, is given by $\min\{k \mid \mathrm{PBT}^{(k)}(\mathcal{C}, \succ) = \mathcal{C}\}$.

Definition 9 (PBTD (Gao et al., 2017)) *Let $\mathcal{C}$ be a concept class. The* preference-based teaching dimension *of $\mathcal{C}$, denoted by* $\mathrm{PBTD}(\mathcal{C})$*, is defined by*

$$\mathrm{PBTD}(\mathcal{C}) = \min\{\mathrm{PBTD}(\mathcal{C}, \succ) \mid \succ \subseteq \mathcal{C} \times \mathcal{C} \text{ and } \succ \text{ forms a strict order on } \mathcal{C}\}\,.$$

The positive preference-based teaching dimension *of $\mathcal{C}$, denoted by* $\mathrm{PBTD}^+(\mathcal{C})$*, is defined analogously to* $\mathrm{PBTD}(\mathcal{C})$*, where the sets S are required to contain only positive examples.*

Note that implicit in the definition of $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ)$ is the existence of a weakly collusion-avoiding teacher mapping T on $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ)$ that has order at most k.

The following property is crucial when relating PBTD and RTD. Its proof, only a slight generalization of the proof by Gao et al. (2017) for the case of finite concept classes (cf. Lemma 7), asserts essentially that any "least preferred" concept must be taught definitively. Recall that a strict order on $\mathcal{C}$ is *well-founded* if for every element of $C \in \mathcal{C}$ the descending chains starting from C are all finite. We say that a strict order on $\mathcal{C}$ is *strongly well-founded* if for every element of $C \in \mathcal{C}$ the descending chains starting from C are all bounded in length by some constant (depending on C).

Proposition 10 (Gao et al. (2017)) *Let $\mathcal{C}$ be any concept class and $\succ$ any binary relation that forms a well-founded strict order over $\mathcal{C}$. Then,* $\mathrm{PBTD}(\mathcal{C}, \succ) \geq \mathrm{DTD}_{\min}(\mathcal{C})$ *and* $\mathrm{PBTD}^+(\mathcal{C}, \succ) \geq \mathrm{DTD}_{\min}{}^+(\mathcal{C})$.

It follows immediately that if $\succ$ is a strongly well-founded strict order, then $\mathrm{PBTD}(\mathcal{C}, \succ) \geq \mathrm{RTD}(\mathcal{C})$.

On the other hand, suppose $\mathcal{C}$ is any concept class and $\succ_k$ is the binary relation on $\mathcal{C}$ in which $C' \succ_k C$ if and only if $C \in \mathrm{RT}_r^{(k)}(\mathcal{C})$ and $C' \in \mathrm{RT}_{r'}^{(k)}(\mathcal{C})$ for some $r' < r$. Then it is straightforward to confirm that (i) $\succ_k$ is a strongly well-founded strict order on $\mathrm{RT}_*^{(k)}(\mathcal{C})$, and (ii) $\mathrm{RT}_*^{(k)}(\mathcal{C}) \subseteq \mathrm{PBT}^{(k)}(\mathcal{C}, \succ_k)$. Hence, if $\mathrm{RTD}(\mathcal{C}) = k$ then $\mathrm{PBTD}(\mathcal{C}, \succ_k) \leq \mathrm{RTD}(\mathcal{C})$ and $\mathrm{PBTD}^+(\mathcal{C}, \succ_k) \leq \mathrm{RTD}_*^+(\mathcal{C})$.

It was shown in Gao et al. (2017) (cf. Corollary 9) that if $\mathcal{C}$ is a finite concept class, then $\mathrm{PBTD}(\mathcal{C})$ and $\mathrm{RTD}(\mathcal{C})$ coincide, as do $\mathrm{PBTD}^+(\mathcal{C})$ and $\mathrm{RTD}^+(\mathcal{C})$. From the remarks above, it follows immediately that the same relationships hold, even if $\mathcal{C}$ is infinite, provided $\mathrm{PBTD}(\mathcal{C}) = \mathrm{PBTD}(\mathcal{C}, \succ)$, for some strongly well-founded strict order $\succ$ on $\mathcal{C}$ (cf. Lemma 8, in Gao et al. (2017)).

Proposition 11 (Gao et al. (2017)) *If* $\mathrm{PBTD}(\mathcal{C}) = \mathrm{PBTD}(\mathcal{C}, \succ)$*, for some strongly well-founded strict order $\succ$ on $\mathcal{C}$, then* $\mathrm{PBTD}(\mathcal{C}) = \mathrm{RTD}(\mathcal{C})$*. (Similarly for* $\mathrm{PBTD}^+(\mathcal{C})$ *and* $\mathrm{RTD}^+(\mathcal{C})$*.)*

Note that the concept class $\mathcal{C}_B$, described in the introduction, has $\mathrm{DTD}(\mathcal{C}_B) = \mathrm{RTD}(\mathcal{C}_B) = \infty$ but $\mathrm{PBTD}^+(\mathcal{C}_B) = 1$. In addition, for the concept classes $\mathcal{C}_{F+}$ and $\mathcal{C}_{G+}$ defined above, it is straightforward to confirm that $\mathrm{PBTD}^+(\mathcal{C}_{F+}) = \mathrm{PBTD}^+(\mathcal{C}_{G+}) = 1$. These illustrate the fact that PBTD^+ can be an arbitrarily small fraction of RTD.

3. The Limit of Weakly Collusion-Avoiding Teaching

As we have already noted, teacher-learner pairs following the classical teaching dimension model, the recursive teaching model, or the preference-based teaching model are all collusion-avoiding

according to the weak (Goldman-Mathias) criterion (cf. Definition 4). Of these models, the classical teaching dimension model is the one imposing the most constraints on the mapping T, followed by recursive teaching, and preference-based teaching in that order. Consequently, the "teaching complexity" among these models is lowest for preference-based teaching; if every concept in a concept class $\mathcal{C}$ can be taught with at most k examples in any of these models, then every concept in $\mathcal{C}$ can be taught with at most k examples in the preference-based model.

One could still argue that the preference-based model is unnecessarily constraining. Preference-based teaching of a concept class $\mathcal{C}$ relies on a preference relation that induces a strict order on $\mathcal{C}$. However, this strict order is used by the learner only after the teaching set has been communicated, since the learner chooses the unique most preferred concept among those *consistent with the set of examples provided by the teacher*. One might consider loosening the constraints by, for example, demanding only that the set of concepts consistent with any chosen teaching set be ordered under the chosen preference relation (rather than requiring acyclic preferences over the whole concept class). In the same vein, one could relax more conditions—every relaxation might result in a more powerful model of teaching satisfying the weakly collusion-avoiding criterion. In fact, we have already seen in Remark 5 that weakly collusion-avoiding teachers of order 1 exist for the powerset $\mathcal{P}_2$ on two elements, while clearly $\text{PBTD}(\mathcal{P}_2) = 2$.

In the following, we will define the provably most powerful model of teaching that is collusion-avoiding as defined in Definition 4.

3.1 Non-clashing Teaching

Before we define this model formally, we introduce a crucial property that was originally proposed by Kuzmin and Warmuth (2007) in the context of unlabeled sample compression.

Definition 12 (Non-Clashing Teacher) *Let $\mathcal{C}$ be a concept class and T be a teacher mapping on $\mathcal{C}$. We say that T is* non-clashing *(on $\mathcal{C}$) if there are no two distinct $C, C' \in \mathcal{C}$ such that both $T(C)$ is consistent with C' and $T(C')$ is consistent with C.*

Prior to Kuzmin and Warmuth (2007), de la Higuera (1997) used the non-clashing property in the context of grammatical inference, with the added constraint that the size of the sample sets (i.e., of the sets $T(C)$ in our definition) be bounded from above by a polynomial in the size of the underlying representation of C, which in de la Higuera's work would be a generator or an acceptor for a formal language. It turns out that, for a teacher mapping T, the non-clashing property coincides with the weak collusion-avoidance criterion:

Theorem 13 *Let $\mathcal{C}$ be a concept class over the instance space $\mathcal{X}$. Let T be a teacher mapping on $\mathcal{C}$. Then the following two conditions are equivalent:*

1. *T is non-clashing on $\mathcal{C}$.*

2. *T satisfies the weak collusion-avoidance criterion on $\mathcal{C}$.*

Proof In essence, the proof is the same as that used by de la Higuera (1997) for a similar statement.

First, suppose T is a non-clashing teacher mapping, and define L as follows. Given any set S of labelled examples as input, L returns an arbitrary concept $C \in \mathcal{C}$ such that $T(C) \subseteq S$ and C is consistent with S. If no such concept exists then L is undefined on S.

Suppose there is some concept $C \in \mathcal{C}$ such that a given set S of labelled examples is consistent with C and contains $T(C)$. We claim that then such C is uniquely determined. For if there were two distinct concepts $C, C' \in \mathcal{C}$ consistent with S such that $T(C) \cup T(C') \subseteq S$, then $T(C')$, being a subset of S, would be consistent with C and, likewise, $T(C)$ would be consistent with C'—in contradiction to the non-clashing property of T. From the definition of L, it then follows that T satisfies the weak collusion-avoidance criterion.

Second, suppose T is a teacher mapping and there is a mapping L witnessing the fact that T satisfies the weak collusion-avoidance criterion, i.e., for all $C \in \mathcal{C}$, we have $L(S) = L(T(C)) = C$ whenever S is consistent with C and contains $T(C)$. To see that T is non-clashing, suppose two concepts $C, C' \in \mathcal{C}$ are both consistent with $T(C) \cup T(C')$. Then $C = L(T(C)) = L(T(C) \cup T(C')) = L(T(C')) = C'$. ∎

Consequently, teaching with non-clashing teacher mappings is, in terms of the worst-case number of examples required, the most efficient model that satisfies the weak collusion-avoidance criterion. Hence we define the notion of no-clash teaching dimension as follows:

Definition 14 (NCTD) *Let $\mathcal{C}$ be a concept class over the instance space $\mathcal{X}$. Let T be a teacher mapping on $\mathcal{C}$ of order k. Denote by $\mathrm{NCT}^{(k)}(\mathcal{C}, T)$ the largest subset $\mathcal{C}' \subset \mathcal{C}$ such that T is non-clashing on $\mathcal{C}'$. The* No-Clash Teaching Dimension *of $\mathcal{C}$ with respect to T, denoted by $\mathrm{NCTD}(\mathcal{C}, T)$, is defined as $\min\{k \mid \mathrm{NCT}^{(k)}(\mathcal{C}, T) = \mathcal{C}\}$. Finally, the* No-Clash Teaching Dimension *of $\mathcal{C}$, denoted $\mathrm{NCTD}(\mathcal{C})$, is defined by*

$$\mathrm{NCTD}(\mathcal{C}) = \min\{\mathrm{NCTD}(\mathcal{C}, T) \mid T \text{ is a non-clashing teacher mapping for } \mathcal{C}\}.$$

It is immediate from Theorem 13 that $\mathrm{NCTD}(\mathcal{C})$ corresponds to the order of the minimum order teaching mapping T that satisfies the weak collusion-avoidance criterion.

As with earlier teaching notions, it is natural to study a variant of non-clashing teaching that uses positive examples only.

Definition 15 (NCTD$^+$) *Let $\mathcal{C}$ be a concept class over the domain $\mathcal{X}$. We define $\mathrm{NCTD}^+(\mathcal{C}) = \min\{\mathrm{ord}(T, \mathcal{C}) \mid T \text{ is a positive non-clashing teacher mapping for } \mathcal{C}\}$.*

Furthermore, for finite concept classes, it will be helpful to have the notion of average no-clash teaching dimension:

Definition 16 (NCTD$_\mathrm{avg}$) *Let $\mathcal{C}$ be a finite concept class. The* Average No-Clash Teaching Dimension *of $\mathcal{C}$, denoted by $\mathrm{NCTD}_\mathrm{avg}(\mathcal{C})$, is defined as*

$$\mathrm{NCTD}_\mathrm{avg}(\mathcal{C}) = \min\left\{\frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} |T(C)| \,\middle|\, T \text{ is a non-clashing teacher mapping for } \mathcal{C}\right\}.$$

Remark 17 It follows immediately from the pigeonhole principle that

$$\mathrm{NCTD}(\mathcal{C}) \geq \lceil \mathrm{NCTD}_\mathrm{avg}(\mathcal{C}) \rceil.$$

3.2 Properties of NCTD and NCTD^+

Since NCTD refers to the optimal teaching complexity among those studied in this paper, and characterizes the complexity of teaching without collusion in the weaker sense (as defined by Goldman and Mathias), we will now turn to a deeper study of this parameter.

3.2.1 LOWER BOUNDS ON NCTD AND NCTD^+

To establish lower bounds on NCTD and NCTD^+ for finite concept classes, we first show that $\mathrm{NCTD}(\mathcal{C})$ must be at least as large as the smallest d satisfying $|\mathcal{C}| \leq 2^d \binom{|\mathcal{X}|}{d}$. This bound actually applies to any teacher mapping that uses teaching sets of size at most d. A similar statement then follows for NCTD^+. In fact, we prove a slightly stronger result, replacing $|\mathcal{X}|$ with a potentially smaller value:

Definition 18 *We define $\mathcal{X}_T \subseteq \mathcal{X}$ as the set of instances that are part of a labelled example in a teaching set $T(C)$ for some $C \in \mathcal{C}$. Moreover, we define*

$$X(\mathcal{C}) = \min\{|\mathcal{X}_T| : \ T \text{ is a non-clashing teacher mapping for } \mathcal{C} \text{ with } \mathrm{ord}(\mathcal{C}, T) = \mathrm{NCTD}(\mathcal{C})\} \,.$$

Intuitively, $X(\mathcal{C})$ is the smallest number of instances that must be employed by any optimal non-clashing teacher mapping for $\mathcal{C}$. Likewise, we define $X^+(\mathcal{C})$ for positive non-clashing teaching.

Theorem 19 *Let $\mathcal{C}$ be any concept class.*

1. *If $\mathrm{NCTD}(\mathcal{C}) = d$, then $|\mathcal{C}| \leq 2^d \binom{X(\mathcal{C})}{d}$.*

2. *If $\mathrm{NCTD}^+(\mathcal{C}) = d$, then $|\mathcal{C}| \leq \sum_{i=0}^{d} \binom{X^+(\mathcal{C})}{i}$.*

Proof To prove statement 1, let $\mathcal{X}'$ be a subset of size $X(\mathcal{C})$ of $\mathcal{X}$. Let $C \mapsto T(C) \subseteq \mathcal{X}' \times \{0, 1\}$ be a consistent and non-clashing mapping which witnesses that $\mathrm{NCTD}(\mathcal{C}) = d$, and let L be the mapping such that $L(T(C)) = C$ for all $C \in \mathcal{C}$. By Proposition 6, one may assume without loss of generality that $|T(C)| = d$ for all $C \in \mathcal{C}$. Since T is an injective mapping and there are only $2^d \binom{X(\mathcal{C})}{d}$ labelled teaching sets at our disposal, the claim follows.

Statement 2 is proven analogously, taking into consideration that, in the NCTD^+ case, we do not have an analogous statement to Proposition 6, since a concept does not in general contain d or more elements. Note that the formula has no factors 2^i since there are no options for labelling the instances in any set $T(C)$. ∎

We will next establish a useful lower bound on $\mathrm{NCTD}(\mathcal{C})$, as well as as a related lower bound on $\mathrm{NCTD}^+(\mathcal{C})$, based on the number of neighbors of any concept in $\mathcal{C}$.

Definition 20 (One-Inclusion Graph (Haussler et al., 1994)) *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$. The one-inclusion graph $G(\mathcal{C})$ is the (undirected) graph given by the vertex set $\mathcal{C}$ and the edge set $\{(C, C') \mid |C \Delta C'| = 1\}$, where $C \Delta C' := (C \setminus C') \cup (C' \setminus C)$. Any edge (C, C') in this graph is labeled with the unique element $x \in \mathcal{X}$ for which $C(x) \neq C'(x)$.*

A concept $C' \in \mathcal{C}$ is a *neighbor* of concept $C \in \mathcal{C}$ if (C, C') is an edge of $G(\mathcal{C})$. If $C' \subseteq \mathcal{C}$ then the *degree* of $C \in \mathcal{C}'$, with respect to $\mathcal{C}'$, denoted as $\deg_{\mathcal{C}'}(C)$, is defined as the number of neighbors of C in the subgraph of $G(\mathcal{C})$ induced on the vertices in $\mathcal{C}'$. The average induced degree of concepts in the subset $\mathcal{C}' \subseteq \mathcal{C}$ is then denoted by

$$\deg_{avg}(\mathcal{C}') := \frac{1}{|\mathcal{C}'|} \cdot \sum_{C \in \mathcal{C}'} \deg_{\mathcal{C}'}(C).$$

The *dominance* of $C \in \mathcal{C}$, denoted as $\mathrm{dom}_{\mathcal{C}}(C)$, is defined as the number of smaller neighbors of C in $\mathcal{C}$, i.e. neighbors that contain exactly one fewer instance than C.

Theorem 21 *Every concept class $\mathcal{C}$ over a finite domain satisfies (i)* $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}) \geq \frac{1}{2} \cdot \deg_{avg}(\mathcal{C})$ *and (ii)* $\mathrm{NCTD}(\mathcal{C}) \geq \max_{\mathcal{C}' \subseteq \mathcal{C}} \lceil \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}') \rceil$.

Proof For assertion (i), let T be any non-clashing teacher mapping for $\mathcal{C}$. If C_1 and C_2, both in $\mathcal{C}'$, are neighbors, and $C_1(x) \neq C_2(x)$, then at least one of the sets $T(C_1), T(C_2)$ must contain x. We obtain $\sum_{C \in \mathcal{C}'} |T(C)| \geq \frac{1}{2} \cdot \sum_{C \in \mathcal{C}'} \deg_{\mathcal{C}'} C) = |\mathcal{C}'| \cdot \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}')$. Assertion (ii) is immediate from (i), by Remark 17. ∎

Theorem 22 *Every concept class $\mathcal{C}$ over a finite domain satisfies*

$$\mathrm{NCTD}^+(\mathcal{C}) \geq \max_{C \in \mathcal{C}} \mathrm{dom}_{\mathcal{C}}(C).$$

Proof If the smaller neighbor C' of $C \in \mathcal{C}$ differs from C on instance x_i, then $(x_i, 1)$ must be used in teaching C. Hence, every $C \in \mathcal{C}$ must have a positive teaching set of size at least $\mathrm{dom}_{\mathcal{C}}(C)$. ∎

Although the lower bounds in Theorems 21 and 22 are not expected to be attained very often, the following Remark shows that they are sometimes tight:

Remark 23 Let $\mathcal{P}_m$ be the powerset over the domain $\{x_1, \ldots, x_m\}$. Since every concept in $\mathcal{P}_m$ has degree m, clearly $\deg_{avg}(\mathcal{P}_m) = m$. It follows from Theorem 21 that $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{P}_m) \geq m/2$ and hence $\mathrm{NCTD}(\mathcal{P}_m) \geq \lceil m/2 \rceil$. Furthermore, since $\mathrm{dom}_{\mathcal{P}_m}(\{x_1, \ldots, x_m\}) = m$, it follows from Theorem 22 that $\mathrm{NCTD}^+(\mathcal{P}_m) \geq m$. But the positive mapping T that maps $S \in \mathcal{P}_m$ to $S \times \{1\}$ is trivially non-clashing, and hence $\mathrm{NCTD}^+(\mathcal{P}_m) = m$ and $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{P}_m) = m/2$. By Remark 5, we have $\mathrm{NCTD}(\mathcal{P}_2) = 1$. As we shall see in Theorem 29, this generalizes to $\mathrm{NCTD}(\mathcal{P}_m) = \lceil m/2 \rceil$.

We note that the maximum degree of a concept in $\mathcal{C}$ is in general not an upper bound on $\mathrm{NCTD}(\mathcal{C})$. For example, if we consider the concept class $\mathcal{C}$ consisting of subsets of size k of some domain of size n, then all concepts in $\mathcal{C}$ have degree zero yet, for n sufficiently large, $\mathrm{NCTD}(\mathcal{C}) = k$ (since for large enough n the size of $\mathcal{C}$ exceeds the number of possible teaching sets in a normal-form teaching mapping T for $\mathcal{C}$ with $\mathrm{ord}(T, \mathcal{C}) < k$.)

The degree lower bounds for $\mathrm{NCTD}_{\mathrm{avg}}$ and NCTD in Theorem 21 can be generalized by taking into consideration the fact that all concept pairs, not just neighbours need to be distinguished by their teaching sets. Specifically, if we construct the complete graph on the set of concepts $\mathcal{C}$ and (i) label

the edge joining concept C and concept C' with $C\Delta C'$, and (ii) label each concept C by its teaching set $T(C)$ then the no-clash condition requires that each edge label $C\Delta C'$ must intersect the union of the labels of its endpoints $T(C) \cup T(C')$.

Thus $\mathrm{NCTD}(\mathcal{C})$ corresponds to the minimum, over all such labelings, of the maximum size vertex label used. We can lower bound $\mathrm{NCTD}_{\mathrm{avg}}$ (and hence NCTD) by observing a lower bound on the *total* size of the vertex labels (teaching sets). One such bound arises by choosing a subset $\mathcal{C}'$ of $\mathcal{C}$ and a subset M of the edges joining vertices in $\mathcal{C}'$. If the labels of the edges in M are all *incidence-disjoint*, meaning they do not intersect the labels of incident edges in M, then each edge in M creates a demand on the labels of its endpoints that cannot be satisfied by a vertex label that is used to satisfy the demand of any other edge in M. Hence,

Theorem 24 *If $\mathcal{C}' \subseteq \mathcal{C}$ and M is any subset of the edges joining vertices in $\mathcal{C}'$ whose associated labels are incidence-disjoint, then* $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}) \geq |M|/|\mathcal{C}'|$ *and (ii)* $\mathrm{NCTD}(\mathcal{C}) \geq \lceil |M|/|\mathcal{C}'| \rceil$.

Note that this generalizes Theorem 21 since the edges joining neighbors have labels that are incidence disjoint. It is easy to construct examples, such as the set $\{a_1^1, a_2^1\} \times \{a_1^2, a_2^2\} \times \ldots \times \{a_1^k, a_2^k\}$ (or the set of all subsets of a fixed size k) for which Theorem 21 gives only a trivial lower bound, yet an optimal lower bound (of k) follows easily from Theorem 24.

3.2.2 SUB-ADDITIVITY OF NCTD AND NCTD$^+$

In this section, we will show that the NCTD is sub-additive with respect to the free combination of concept classes. As an application of this result, we will determine the NCTD of the powerset over any finite domain $\mathcal{X}$. While the powerset is a rather special concept class, knowing its NCTD will turn out useful to obtain a variety of further results.

Definition 25 (Free Combination of Concept Classes) *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be concept classes over disjoint domains $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Then the* free combination $\mathcal{C}_1 \sqcup \mathcal{C}_2$ *of $\mathcal{C}_1$ and $\mathcal{C}_2$ is a concept class over the domain $\mathcal{X}_1 \cup \mathcal{X}_2$ defined by $\mathcal{C}_1 \sqcup \mathcal{C}_2 = \{C_1 \cup C_2 | \ C_1 \in \mathcal{C}_1$ and $C_2 \in \mathcal{C}_2\}$.*

Lemma 26 *Let $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2$ be the free combination of $\mathcal{C}_1$ and $\mathcal{C}_2$. Moreover, for $i = 1, 2$, let T_i be a non-clashing mapping for $\mathcal{C}_i$. Then, for $T(\mathcal{C}_1 \sqcup \mathcal{C}_2)$ defined by setting $T(C_1 \cup C_2) = T_1(C_1) \cup T_2(C_2)$, we have that T is a non-clashing teacher mapping for $\mathcal{C}_1 \sqcup \mathcal{C}_2$. Moreover, as witnessed by T, NCTD acts sub-additively on $\sqcup$, i.e.,*

$$\mathrm{NCTD}(\mathcal{C}_1 \sqcup \mathcal{C}_2) \leq \mathrm{NCTD}(\mathcal{C}_1) + \mathrm{NCTD}(\mathcal{C}_2). \tag{1}$$

Proof Suppose that concepts $C_{i_1}, C_{j_1} \in \mathcal{C}_1$ and $C_{i_2}, C_{j_2} \in \mathcal{C}_2$ give rise to distinct concepts $C_{i_1} \cup C_{i_2}$ and $C_{j_1} \cup C_{j_2} \in \mathcal{C}_1 \sqcup \mathcal{C}_2$ that clash under T. (Without loss of generality we can assume that $i_1 \neq j_1$.) Then $C_{j_1} \cup C_{j_2}$ is consistent with $T_1(C_{i_1}) \cup T_2(C_{i_2})$ and $C_{i_1} \cup C_{i_2}$ is consistent with $T_1(C_{j_1}) \cup T_2(C_{j_2})$. Hence C_{j_1} is consistent with $T_1(C_{i_1})$ and C_{i_1} is consistent with $T_1(C_{j_1})$, that is concepts C_{i_1} and C_{j_1} in $\mathcal{C}_1$ clash under the mapping T_1. ∎

As we shall see NCTD sometimes acts strictly sub-additively on $\sqcup$; in particular, the composition of optimal mappings for $\mathcal{C}_1$ and $\mathcal{C}_2$ is not necessarily an optimal mapping for $\mathcal{C}_1 \sqcup \mathcal{C}_2$. In contrast, ANCTD acts additively on $\sqcup$:

Lemma 27 *Let $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2$ be the free combination of $\mathcal{C}_1$ and $\mathcal{C}_2$. Moreover, for $i = 1, 2$, let T_i be a non-clashing mapping for $\mathcal{C}_i$. Then, for $T(\mathcal{C}_1 \sqcup \mathcal{C}_2)$ defined by setting $T(C_1 \cup C_2) = T_1(C_1) \cup T_2(C_2)$, we have that T is a non-clashing teacher mapping for $\mathcal{C}_1 \sqcup \mathcal{C}_2$. Moreover, as witnessed by T, $\mathrm{NCTD}_{\mathrm{avg}}$ acts additively on $\sqcup$, i.e.,*

$$\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1 \sqcup \mathcal{C}_2) = \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1) + \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_2). \tag{2}$$

Proof The proof of Lemma 26 above shows that ANCTD acts sub-additively on $\sqcup$, that is $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1 \sqcup \mathcal{C}_2) \leq \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1) + \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_2)$. It remains to show that

$$\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1 \sqcup \mathcal{C}_2) \geq \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1) + \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_2). \tag{3}$$

To this end, let $\mathcal{X}_1$ (resp. $\mathcal{X}_2$) be the domain of concept class $\mathcal{C}_1$ (resp. $\mathcal{C}_2$) and suppose that T is a non-clashing teacher mapping on $\mathcal{C}$ such that $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} |T(C)|$. The following calculation makes use of the fact, for every fixed choice of $C_2 \in \mathcal{C}_2$, the mapping $C_1 \mapsto T(C_1 \cup C_2) \cap \mathcal{X}_1$ is a non-clashing teacher mapping on $\mathcal{C}_1$ (and an analogous remark holds, for reasons of symmetry, when the roles of $\mathcal{C}_1$ and $\mathcal{C}_2$ are exchanged):

$$
\begin{aligned}
\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}) &= \frac{1}{|\mathcal{C}_1| \cdot |\mathcal{C}_2|} \sum_{C_1 \in \mathcal{C}_1} \sum_{C_2 \in \mathcal{C}_2} |T(C_1 \cup C_2|) \\
&= \frac{1}{|\mathcal{C}_2|} \sum_{C \in \mathcal{C}_2} \left(\frac{1}{|\mathcal{C}_1|} \sum_{C \in \mathcal{C}_1} |T(C_1 \cup C_2) \cap \mathcal{X}_1| \right) \\
&\quad + \frac{1}{|\mathcal{C}_1|} \sum_{C \in \mathcal{C}_1} \left(\frac{1}{|\mathcal{C}_2|} \sum_{C \in \mathcal{C}_2} |T(C_1 \cup C_2) \cap \mathcal{X}_2| \right) \\
&\geq \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_1) + \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}_2).
\end{aligned}
$$

■

Remark 28 In Lemma 26, if T_1 and T_2 are positive non-clashing mappings, then the same proof shows that T (a positive non-clashing mapping) witnesses the fact that NCTD^+ also acts sub-additively on $\sqcup$, i.e.,

$$\mathrm{NCTD}^+(\mathcal{C}_1 \sqcup \mathcal{C}_2) \leq \mathrm{NCTD}^+(\mathcal{C}_1) + \mathrm{NCTD}^+(\mathcal{C}_2). \tag{4}$$

Furthermore, since $\sqcup$ is associative, it follows immediately that, for any concept class $\mathcal{C}$, if $\mathcal{C}^k := \mathcal{C}_1 \sqcup \ldots \sqcup \mathcal{C}_k$, where $\mathcal{C}_i := \{C \times \{i\} \mid C \in \mathcal{C}\}$ for $i = 1, \ldots, k$, then

$$\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}^k) = k \cdot \mathrm{NCTD}_{\mathrm{avg}}(\mathcal{C}) \tag{5}$$

and

$$\mathrm{NCTD}(\mathcal{C}^k) \leq k \cdot \mathrm{NCTD}(\mathcal{C}) \text{ and } \mathrm{NCTD}^+(\mathcal{C}^k) \leq k \cdot \mathrm{NCTD}^+(\mathcal{C}). \tag{6}$$

We have already seen, in Remark 23, that $\mathrm{NCTD}_{\mathrm{avg}}(\mathcal{P}_m) = m/2$, $\mathrm{NCTD}^+(\mathcal{P}_m) = m$ and $\mathrm{NCTD}(\mathcal{P}_m) \geq \lceil m/2 \rceil$, where $\mathcal{P}_m$ denotes the powerset over the domain $\{x_1, \ldots, x_m\}$. The sub-additivity results above can be applied in order to determine $\mathrm{NCTD}(\mathcal{P}_m)$ exactly as well.

Theorem 29 *Let $\mathcal{P}_m$ be the powerset over the domain $\{x_1, \ldots, x_m\}$. Then $\mathrm{NCTD}(\mathcal{P}_m) = \lceil m/2 \rceil$.*

Proof It remains to show that $\mathrm{NCTD}(\mathcal{P}_m) \leq \lceil m/2 \rceil$. It suffices to verify this upper bound for even m. But, when m is even, $\mathrm{NCTD}(\mathcal{P}_m) = \mathrm{NCTD}(\mathcal{P}_2^{m/2}) \leq m/2$ follows from (6) and the fact that $\mathrm{NCTD}(\mathcal{P}_2) = 1$ (cf. Remark 23). ∎

Since the NCTD of any concept class over a domain $\mathcal{X}$ is trivially upper bounded by the NCTD of the powerset over $\mathcal{X}$, this result in particular implies that $\lceil |\mathcal{X}|/2 \rceil$ is an upper bound on the NCTD of any concept class over a domain $\mathcal{X}$.

A further consequence of Theorem 29 is that NCTD is sometimes strictly subadditive with respect to the operation of free combination, i.e., that inequality (1) is sometimes strict. An example for that is the free combination $\mathcal{P}_m \sqcup \mathcal{P}_m$ of two copies of $\mathcal{P}_m$ for odd m. Since the domain of $\mathcal{P}_m \sqcup \mathcal{P}_m$ has size $2m$, we obtain $\mathrm{NCTD}(\mathcal{P}_m \sqcup \mathcal{P}_m) = m$, while $\mathrm{NCTD}(\mathcal{P}_m) + \mathrm{NCTD}(\mathcal{P}_m) = 2\lceil \frac{m}{2} \rceil = 2\frac{m+1}{2} = m + 1$.

Another situation (that we will exploit later) where NCTD^+ acts strictly additively on $\sqcup$, is captured in the following:

Lemma 30 *Let $\mathcal{P}_m$ be the powerset over the domain $\{x_1, \ldots, x_m\}$ and let $\mathcal{C}$ be a concept class with domain $\mathcal{X}$ disjoint from $\{x_1, \ldots, x_m\}$. Then,*

$$\mathrm{NCTD}^+(\mathcal{P}_m \sqcup \mathcal{C}) = \mathrm{NCTD}^+(\mathcal{P}_m) + \mathrm{NCTD}^+(\mathcal{C}).$$

Proof By (4) it suffices to show that $\mathrm{NCTD}^+(\mathcal{P}_m \sqcup \mathcal{C}) \geq \mathrm{NCTD}^+(\mathcal{P}_m) + \mathrm{NCTD}^+(\mathcal{C})$. Theorem 22 implies that, for each $C_i \in \mathcal{C}$, any positive non-clashing mapping T for $\mathcal{P}_m \sqcup \mathcal{C}$ must use $m = \mathrm{NCTD}^+(\mathcal{P}_m)$ examples from $\{x_1, \ldots, x_m\}$ to teach the single concept $\{x_1, \ldots, x_m\} \sqcup C_i$ within the concept class $\mathcal{P}_m \sqcup C_i$. So the only way that T could use fewer than $m + \mathrm{NCTD}^+(\mathcal{C})$ examples in total for each concept in $\{x_1, \ldots, x_m\} \sqcup \mathcal{C}$ is if each such concept is taught with exactly m examples from $\{x_1, \ldots, x_m\}$, and hence fewer than $\mathrm{NCTD}^+(\mathcal{C})$ examples from $\mathcal{X}$, a contradiction. ∎

Furthermore, it is easily seen that the average degree acts additively on $\sqcup$:

Lemma 31 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be concept classes over disjoint and finite domains. Then the following holds:*

$$\deg_{avg}(\mathcal{C}_1 \sqcup \mathcal{C}_2) = \deg_{avg}(\mathcal{C}_1) + \deg_{avg}(\mathcal{C}_2). \tag{7}$$

Proof Let $\mathcal{C} := \mathcal{C}_1 \sqcup \mathcal{C}_2$. The concepts in $\mathcal{C}$ that are neighbors of $C_1 \cup C_2 \in \mathcal{C}$ are precisely the concepts of the form $C_1 \cup C_2'$ or $C_1' \cup C_2$ where C_2' is a neighbor of C_2 in $\mathcal{C}_2$ and C_1' is a neighbor of C_1 in $\mathcal{C}_1$. Hence

$$\deg_{\mathcal{C}}(C_1 \cup C_2) = \deg_{\mathcal{C}_1}(C_1) + \deg_{\mathcal{C}_2}(C_2).$$

Moreover $|\mathcal{C}| = |\mathcal{C}_1| \cdot |\mathcal{C}_2|$. It follows that

$$\sum_{C \in \mathcal{C}} \deg_{\mathcal{C}}(C) = \sum_{C_1 \in \mathcal{C}_1} \sum_{C_2 \in \mathcal{C}_2} \deg_{\mathcal{C}}(C_1 \cup C_2) = |\mathcal{C}_2| \cdot \sum_{C_1 \in \mathcal{C}_1} \deg_{\mathcal{C}_1}(C_1) + |\mathcal{C}_1| \cdot \sum_{C_2 \in \mathcal{C}_2} \deg_{\mathcal{C}_2}(C_2).$$

Division by $|\mathcal{C}_1| \cdot |\mathcal{C}_2|$ immediately yields (7). ∎

The free combination of classes with a tight degree lower bound is again a class with a tight degree lower bound:

Corollary 32 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two concept classes over disjoint and finite domains, and let $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2$. Then $\mathrm{NCTD}(\mathcal{C}_i) = \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}_i)$ for $i = 1, 2$ implies that $\mathrm{NCTD}(\mathcal{C}) = \frac{1}{2} \cdot \deg_{avg}(\mathcal{C})$.*

Proof The assertion is evident from the chain of inequalities:

$$\mathrm{NCTD}(\mathcal{C}) \overset{(1)}{\leq} \mathrm{NCTD}(\mathcal{C}_1) + \mathrm{NCTD}(\mathcal{C}_2) = \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}_1) + \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}_2) \overset{(7)}{=} \frac{1}{2} \cdot \deg_{avg}(\mathcal{C})$$

and Theorem 21. ∎

4. Relation Between NCTD and Other Learning-theoretic Parameters

In this section, we set NCTD in relation to PBTD and VCD, as well as to the smallest possible size of a sample compression scheme for a given concept class.

While most of the statements made in this section are immediate consequences of known results on PBTD and VCD, we consider it worthwhile gathering these results for the sake of providing a complete picture of how NCTD relates to other important complexity parameters. Moreover, we provide non-trivial new results towards the so far unresolved question of how large the gap between NCTD and PBTD can be.

4.1 PBTD

While we have already seen that

$$\mathrm{NCTD}(\mathcal{C}) \leq \mathrm{PBTD}(\mathcal{C}) \text{ and } \mathrm{NCTD}^+(\mathcal{C}) \leq \mathrm{PBTD}^+(\mathcal{C}),$$

for any concept class $\mathcal{C}$, it is far more intricate to determine the ratios by which the individual parameters in these inequalities can differ.

4.1.1 EXAMPLES OF CONCEPT CLASSES WITH $\mathrm{NCTD} < \mathrm{PBTD}$

The inequality $\mathrm{NCTD}(\mathcal{C}) \leq \mathrm{PBTD}(\mathcal{C})$ is sometimes strict, as witnessed by Theorem 29, which states that $\mathrm{NCTD}(\mathcal{P}_m) = \lceil m/2 \rceil$. By comparison, $\mathrm{PBTD}(\mathcal{P}_m) = m$. In particular, this yields a family of concept classes of strictly increasing NCTD for which PBTD exceeds NCTD by a factor of 2. The fact that the inequality $\mathrm{NCTD}^+(\mathcal{C}) \leq \mathrm{PBTD}^+(\mathcal{C})$ is sometimes strict is witnessed by the simple class $\mathcal{C}_A$ described in the introduction, with $\mathrm{NCTD}^+(\mathcal{C}_A) = 1$. Since no concept in $\mathcal{C}_A$ has a positive teaching set of size 1, Proposition 10 implies $\mathrm{PBTD}^+(\mathcal{C}_A) = 2$. In particular, these examples witness that Proposition 10 does *not* hold for non-clashing teaching.

Appendix A shows two interesting examples of concept classes for which NCTD^+ equals 1 while PBTD equals 3. The main result of this subsection is that Sylvester-Hadamard matrices can be used to define a natural family of finite concept classes, for which PBTD and DTD_{min} exceed NCTD by a factor of at least 4.

Example 1 *In this example, as usual, we associate a binary matrix with a concept class in which the rows are concepts and the columns are instances. Sylvester's standard construction of Hadamard matrices (with zeroes replacing the -1 entries) can then be used to construct concept classes of interest to our study.*

Define the 1×1-matrix H_0 to consist of only the entry 1. In general, for $n \in \mathbb{N} \setminus \{0\}$, define the $2^n \times 2^n$-matrix H_n as

$$\begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & \overline{H}_{n-1} \end{pmatrix} ,$$

where $\overline{H}$ results from H by flipping every 1 to a 0 and vice versa.

It is well known that $H_{m+n} = H_m \otimes H_n$, where $\otimes$ denotes the Kronecker product (where 0 operates as -1).

The recursive structure of H_n, $n > 0$, makes possible a straightforward inductive proof of the following:

Observation 33 *If we consider the 2^{n-1} rows Y_i of H_n that contain a fixed value (0 or 1) in column $i > 1$ then there is an associated set of 2^{n-1} columns X_i such that the entries belonging to rows Y_i and columns X_i form either H_{n-1} or $\overline{H}_{n-1}$. It follows immediately from this that if we fix the value of any $s < n$ instances then provided there is at least one consistent concept there must be at least 2^{n-s} consistent concepts. Hence, $\mathrm{DTD}_{min}(H_n) = n$.*

Note that the rows of a Sylvester-Hadamard matrix are called Walsh functions, and are an equivalent representation of parity functions. In particular, the concept class consisting of the rows of H_n is equivalent to the class Π_n of parity functions over the domain $X_n = \{0, 1\}^n$, i.e., $\Pi_n = \{\chi_w \mid w \in \{0, 1\}^n\}$ where, for every $x \in \{0, 1\}^n$,

$$\chi_w(x) = \langle w, x \rangle = w_1 x_1 \oplus \ldots \oplus w_n x_n \,.$$

Here $\oplus$ denotes addition modulo 2 and $\langle \cdot, \cdot \rangle$ denotes the inner product modulo 2. This equivalence can be exploited to show that, for all $n \geq 1$, we have $\mathrm{NCTD}(H_{4n}) \leq n$.

Theorem 34 *For all $n \geq 1$, we have $\mathrm{NCTD}(H_{4n}) \leq n$.*

Proof We will show that $\mathrm{NCTD}(\Pi_{4n}) \leq n$.

Fix a teaching mapping

$$\chi_{\boldsymbol{u}} \mapsto \{(\tilde{\boldsymbol{u}}, \chi_{\boldsymbol{u}}(\tilde{\boldsymbol{u}}))\} \tag{8}$$

which witnesses that $\mathrm{NCTD}(\Pi_4) = 1$. Such a mapping exists; an example is given in Table 1.

A vector $\bar{\boldsymbol{u}} \in X_{4n}$ can be decomposed into vectors $\boldsymbol{u_1}, \ldots, \boldsymbol{u_n} \in X_4$. Let T be the teacher mapping which assigns to $\chi_{\bar{\boldsymbol{u}}}$ the instances

$$(\tilde{\boldsymbol{u}_1}, \boldsymbol{0}, \ldots, \boldsymbol{0}), \ldots, (\boldsymbol{0}, \ldots, \boldsymbol{0}, \tilde{\boldsymbol{u}_n})$$

together with their $\chi_{\bar{\boldsymbol{u}}}$-labels. Consider another parity function $\chi_{\bar{\boldsymbol{v}}} \neq \chi_{\bar{\boldsymbol{u}}}$ and its teaching set consisting of the instances

$$(\tilde{\boldsymbol{v}_1}, \boldsymbol{0}, \ldots, \boldsymbol{0}), \ldots, (\boldsymbol{0}, \ldots, \boldsymbol{0}, \tilde{\boldsymbol{v}_n})$$

and their $\chi_{\bar{v}}$-labels. Suppose that $\chi_{\bar{v}}$ is consistent with the teaching set of $\chi_{\bar{u}}$. Fix some $j \in [n]$ such that $\boldsymbol{u}_j \neq \boldsymbol{v}_j$. It follows that

$$\chi_{\boldsymbol{v}_j}(\tilde{\boldsymbol{u}}_j) = \chi_{\bar{v}}(\boldsymbol{0}, \ldots, \boldsymbol{0}, \tilde{\boldsymbol{u}}_j, \boldsymbol{0}, \ldots, \boldsymbol{0}) = \chi_{\bar{u}}(\boldsymbol{0}, \ldots, \boldsymbol{0}, \tilde{\boldsymbol{u}}_j, \boldsymbol{0}, \ldots, \boldsymbol{0}) = \chi_{\boldsymbol{u}_j}(\tilde{\boldsymbol{u}}_j).$$

Since (8) is a teacher mapping that witnesses $\mathrm{NCTD}(\Pi_4) = 1$, it follows that

$$\chi_{\boldsymbol{u}_j}(\tilde{\boldsymbol{v}}_j) \neq \chi_{\boldsymbol{v}_j}(\tilde{\boldsymbol{v}}_j).$$

Hence

$$\chi_{\bar{u}}(\boldsymbol{0}, \ldots, \boldsymbol{0}, \tilde{\boldsymbol{v}}_j, \boldsymbol{0}, \ldots, \boldsymbol{0}) = \chi_{\boldsymbol{u}_j}(\tilde{\boldsymbol{v}}_j) \neq \chi_{\boldsymbol{v}_j}(\tilde{\boldsymbol{v}}_j) = \chi_{\bar{v}}(\boldsymbol{0}, \ldots, \boldsymbol{0}, \tilde{\boldsymbol{v}}_j, \boldsymbol{0}, \ldots, \boldsymbol{0}),$$

which shows that $\chi_{\bar{u}}$ is not consistent with the teaching set of $\chi_{\bar{v}}$. Thus T is a valid teacher mapping of order n in the no-clash model, which concludes the proof. ∎

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
C_1	1	1	1	1	1	1	1	1	**1**	1	1	1	1	1	1	1
C_2	1	0	1	0	1	0	1	0	1	**0**	1	0	1	0	1	0
C_3	1	1	0	0	1	1	0	0	1	1	0	**0**	1	1	0	0
C_4	1	0	0	1	1	0	0	1	1	0	0	1	1	0	**0**	1
C_5	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	**0**
C_6	1	0	1	0	0	1	0	1	1	0	1	0	**0**	1	0	1
C_7	1	1	0	0	0	0	1	1	1	1	0	0	0	**0**	1	1
C_8	1	0	0	1	0	1	1	0	1	0	**0**	1	0	1	1	0
C_9	1	1	1	1	1	1	1	1	**0**	0	0	0	0	0	0	0
C_{10}	1	0	1	0	1	0	1	0	0	**1**	0	1	0	1	0	1
C_{11}	1	1	0	0	1	1	0	0	0	0	1	**1**	0	0	1	1
C_{12}	1	0	0	1	1	0	0	1	0	1	1	0	0	1	**1**	0
C_{13}	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	**1**
C_{14}	1	0	1	0	0	1	0	1	0	1	0	1	**1**	0	1	0
C_{15}	1	1	0	0	0	0	1	1	0	0	1	1	1	**1**	0	0
C_{16}	1	0	0	1	0	1	1	0	0	1	**1**	0	1	0	0	1

Table 1: The concept class H_4 ($= \Pi_4$) with a non-clashing teacher mapping of order 1 shown with highlighted entries. Note that the subset $\{x_9 \ldots, x_{16}\}$ of domain instances is already sufficient for defining this non-clashing teacher. In particular, the "right half" of the matrix H_4 yields a concept class of 16 instances over a domain of size 8 for which PBTD equals 4 while NCTD equals 1.

Corollary 35 *For all $n \in \mathbb{N}$, we have $\mathrm{PBTD}(H_n) = n$, while $\mathrm{NCTD}(H_{4n}) \leq n$.*

Recently, a family of concept classes was constructed for which the gap between NCTD and PBTD becomes unbounded (Simon, 2022). However, the natural examples provided here may provide further insights into the intricate relationship between these two parameters under certain constraints on the structure of the underlying concept class.

4.1.2 SHORTEST-PATH-CLOSED CONCEPT CLASSES

The notion of one-inclusion graph (Haussler et al., 1994) is helpful in studying teaching complexity parameters, since, for any so-called *shortest-path-closed concept class* $\mathcal{C}$, the one-inclusion graph provides a simple representation of all smallest definitive teaching sets of any concept in $\mathcal{C}$.

Recall from Definition 20 that the one-inclusion graph $G(\mathcal{C})$ of a concept class $\mathcal{C}$ is the (undirected) graph given by the vertex set $\mathcal{C}$ and the edge set $\{(C, C') \mid |C\Delta C'| = 1\}$, where $C\Delta C' := (C \setminus C') \cup (C' \setminus C)$.

Definition 36 (Shortest-Path-Closed Concept Class) *Suppose that edge (C, C') in $G(\mathcal{C})$ is labeled with the unique element $x \in \mathcal{X}$ for which $C(x) \neq C'(x)$. The concept class $\mathcal{C}$ is called shortest-path-closed if, for any k and any two concepts $C, C' \in \mathcal{C}$ with $|C\Delta C'| = k$, there is a path of length k in $G(\mathcal{C})$ from C to C'.*

It is obvious that $\mathcal{P}_m$, the powerset over the domain $\{x_1, \ldots, x_m\}$, is shortest-path-closed.

As noted by Doliwa et al. (2014), in a shortest-path-closed class, the set of edge labels incident to any concept C corresponds to the unique smallest definitive teaching set of C. Now consider a non-clashing teacher mapping T on a shortest-path-closed class $\mathcal{C}$. Since no two concepts can be consistent with one another's teaching sets, for every edge (C, C') in $G(\mathcal{C})$, at least one of the concepts C, C' must use the label of (C, C') in its teaching set under T.

The following simple characterization of shortest-path-closed classes whose NCTD equals 1 follows directly:

Theorem 37 *Let $\mathcal{C}$ be shortest-path-closed. Then the following statements are equivalent:*

1. *$\mathrm{NCTD}(\mathcal{C}) = 1$;*

2. *The one-inclusion graph $G(\mathcal{C})$ corresponding to $\mathcal{C}$ has at most one cycle.*

Proof (Sketch) If $G(\mathcal{C})$ has more than one cycle then, following iterated removal of vertices of degree one, which preserves the property of being shortest path closed, the resulting graph has minimum degree 2. Since the resulting graph has average degree greater than 2 (otherwise, since it must be connected, it is a simple cycle), it follows, by Theorem 21, that $\mathrm{NCTD}(\mathcal{C}) > 1$.

If $G(\mathcal{C})$ is acyclic, then it is a tree; this implies $\mathrm{RTD}(\mathcal{C}) = 1$ (Doliwa et al., 2014), and thus clearly $\mathrm{NCTD}(\mathcal{C}) = 1$.

If $G(\mathcal{C})$ has exactly one cycle, then this cycle has even length, and each edge label occurring on the cycle occurs exactly twice. Shortest-path-closedness further implies that the two occurrences of any edge label on the cycle are directly opposite to one another, i.e., there are instances $x_1, \ldots, x_k \in \mathcal{X}$ and a vertex C_0 on the cycle, such that the edge labels following the cycle in one direction, starting from C_0, occur in order $x_1, \ldots, x_k, x_1, \ldots, x_k$. Note that any vertex on the cycle may also be the root of a subtree of $G(\mathcal{C})$. Shortest-path-closedness can again be used to show that (i) none of the instances $x_1, \ldots, x_k$ occur as edge labels in any of these subtrees; and (ii) every instance $x \in \mathcal{X} \setminus \{x_1, \ldots, x_k\}$ occurs at most once as an edge label in $G(\mathcal{C})$. Now a non-clashing teacher mapping of order 1 can be defined as follows. First, all edges on the cycle are directed away from C_0 in the direction of the edge labeled x_1 incident to C_0. Then C_0 is taught with $(x_1, C_0(x_1))$, and subsequent concepts on the directed cycle are taught with the instance of their outgoing edge, labeled accordingly. (Clearly, the two occurrences of any instance x_i, for $1 \leq i \leq k$, occur with

opposite binary labels in the corresponding teaching sets.) Second, every non-root concept C in a subtree is taught using only the pair $(x, C(x))$, where x is the label of the edge pointing from C in the direction of the root of the tree containing C. It is easy to verify that this teacher mapping is non-clashing. Thus $\text{NCTD}(\mathcal{C}) = 1$. ∎

Corollary 38 *Let $\mathcal{C}$ be a shortest-path-closed concept class. If $\text{NCTD}(\mathcal{C}) = 1$, then $\text{RTD}(\mathcal{C}) \leq 2$.*

Proof (Sketch.) We know from Theorem 37 that $G(\mathcal{C})$ has at most one cycle. If $G(\mathcal{C})$ has no cycle, then a result by Doliwa et al. (2014) implies that $\text{RTD}(\mathcal{C}) = 1$. If $G(\mathcal{C})$ has a cycle, then each vertex not on that cycle belongs to a tree; using the same result by Doliwa et al. (2014), we can assign (recursive) teaching sets of size 1 to all non-root vertices of any sub-tree of $G(\mathcal{C})$. Finally, only the vertices on the cycle remain. These all have degree 2, i.e., definitive teaching sets of size 2 with respect to the class of all concepts on the cycle. Thus $\text{RTD}(\mathcal{C}) = 2$. ∎

4.2 VCD

It is well known that for any $k \in \mathbb{N}$, $k \geq 1$, there exists a finite concept class $\mathcal{C}$ such that $\text{DTD}^+(\mathcal{C}) = \text{DTD}(\mathcal{C}) = 1$ and $\text{VCD}(\mathcal{C}) = k$ (Goldman and Kearns, 1995). Hence, trivially, VCD cannot be upper-bounded by a function that depends only on NCTD.

Hu et al. (2017) proved that, when $\text{VCD}(\mathcal{C}) = d$, the recursive teaching dimension of $\mathcal{C}$ is no larger than $39.3752 \cdot d^2 - 3.6330 \cdot d$. Obviously, the same upper bound applies to NCTD, i.e., $\text{NCTD}(\mathcal{C})$ is upper-bounded by a function quadratic in $\text{VCD}(\mathcal{C})$.

So far, there is no concept class for which NCTD is known to exceed VCD. Note that any such concept class would have to fulfill PBTD > VCD as well. We tested those classes for which PBTD > VCD is known from the literature, but found that all of them satisfy NCTD $\leq$ VCD.

As an example, here we present "Warmuth's class." This concept class, shown in Table 2, was communicated by Manfred Warmuth and proven by Darnstädt et al. (2016) to be the smallest concept class for which PBTD exceeds VCD. In particular, $\text{VCD}(\mathcal{C}_W) = 2$ while $\text{PBTD}(\mathcal{C}_W) = 3$.

	x_1	x_2	x_3	x_4	x_5		x_1	x_2	x_3	x_4	x_5
C_1	**1**	0	0	0	**1**	C_1'	**1**	0	**1**	0	**1**
C_2	**1**	**1**	0	0	0	C_2'	1	**1**	0	**1**	0
C_3	0	**1**	**1**	0	0	C_3'	0	1	**1**	0	**1**
C_4	0	0	**1**	**1**	0	C_4'	**1**	0	**1**	**1**	0
C_5	0	0	0	**1**	**1**	C_5'	0	**1**	0	**1**	**1**

Table 2: Warmuth's class $\mathcal{C}_W$, with the highlighted entries (in bold) corresponding to the images of a positive non-clashing teacher mapping. The domain of this class is $\{x_1, \ldots, x_5\}$, and it contains 10 concepts, named C_1 through C_5 and C_1' through C_5'.

Proposition 39 $\text{NCTD}(\mathcal{C}_W) = \text{NCTD}^+(\mathcal{C}_W) = 2$.

Proof The highlighted labels in Table 2 correspond to a positive non-clashing mapping for $\mathcal{C}_W$, which immediately shows that $\mathrm{NCTD}^+(\mathcal{C}_W) \leq 2$ and thus $\mathrm{NCTD}(\mathcal{C}_W) \leq 2$. To show that $\mathrm{NCTD}(\mathcal{C}_W) \geq 2$, suppose by way of contradiction that $\mathrm{NCTD}(\mathcal{C}_W) = 1$. Then there is a non-clashing teacher mapping T that assigns every concept in $\mathcal{C}_W$ a teaching set of size 1.

Since C_1 and C_1' differ only on the instance x_3, the mapping T must fulfill either $T(C_1) = \{(x_3, 0)\}$ or $T(C_1') = \{(x_3, 1)\}$.

Case 1. $T(C_1) = \{(x_3, 0)\}$. Since C_2 is consistent with $T(C_1)$, the teaching set for C_2 must be inconsistent with C_1. In particular, $T(C_2) \neq \{(x_4, 0)\}$. This implies $T(C_2') = \{(x_4, 1)\}$, since x_4 is the only instance on which C_2 and C_2' disagree. By an analogous argument concerning C_5 and C_5', one obtains $T(C_5') = \{(x_2, 1)\}$. Now T has a clash on C_2' and C_5', which is a contradiction.

Case 2. $T(C_1') = \{(x_3, 1)\}$. One argues as in Case 1, with C_3' and C_4' in place of C_2 and C_5, yielding $T(C_3) = \{(x_5, 0)\}$ and $T(C_4) = \{(x_1, 0)\}$. This is a clash, resulting in a contradiction.

As both cases result in a contradiction, we have $\mathrm{NCTD}(\mathcal{C}_W) > 1$ and thus $\mathrm{NCTD}(\mathcal{C}_W) = 2$. Since NCTD^+ is an upper bound on NCTD, we also have $\mathrm{NCTD}^+(\mathcal{C}_W) = 2$. ■

While the general relationship between the parameters NCTD and VCD remains open, it turns out that $\mathrm{NCTD}(\mathcal{C})$ is upper-bounded by $\mathrm{VCD}(\mathcal{C})$ when $\mathcal{C}$ is a finite *maximum* class. For a finite instance space $\mathcal{X}$, a concept class $\mathcal{C}$ of VC dimension d is called maximum if its size $|\mathcal{C}|$ meets Sauer's upper bound $\sum_{i=0}^{d} \binom{|\mathcal{X}|}{i}$ (Sauer, 1972) with equality. Recently, Chalopin et al. (2022) showed that every finite maximum class $\mathcal{C}$ admits a so-called *representation map*, i.e., a function r that maps every concept in $\mathcal{C}$ to a set of at most $d(= \mathrm{VCD}(\mathcal{C}))$ instances, in a way that no two distinct concepts $C, C' \in \mathcal{C}$ both agree on all the instances in $r(C) \cup r(C')$. By definition, any representation map is, translated into our setting, simply a non-clashing teacher mapping of order d for $\mathcal{C}$. Therefore, the result by Chalopin et al. implies that $\mathrm{NCTD}(\mathcal{C}) \leq \mathrm{VCD}(\mathcal{C})$ for finite maximum $\mathcal{C}$.

4.3 A Note on Sample Compression

Intuitively, a sample compression scheme (Littlestone and Warmuth, 1986) for a (possibly infinite) concept class $\mathcal{C}$ provides a lossless compression of every set S of labeled examples for any concept in $\mathcal{C}$ in the form of a subset of S. It was proven that the existence of a finite upper bound on the size of the compression sets is equivalent to PAC-learnability, i.e., to finite VC-dimension (Moran and Yehudayoff, 2016; Littlestone and Warmuth, 1986). Open for over 30 years now is the question how closely such an upper bound can be related to the VC-dimension.

Formally, a sample compression scheme of size k for a concept class $\mathcal{C}$ over $\mathcal{X}$ is a pair (f, g) of mappings, where, for every sample set S consistent with some concept $C \in \mathcal{C}$, (i) f maps S to a subset $f(S) \subseteq S$ with $|f(S)| \leq k$; and (ii) $g(f(S))$ maps the compressed set to a concept C' over $\mathcal{X}$ (not necessarily in $\mathcal{C}$) that is consistent with S. By $\mathrm{CN}(\mathcal{C})$ we denote the size of the smallest-size sample compression scheme for $\mathcal{C}$. The open question then is whether $\mathrm{CN}(\mathcal{C})$ is upper-bounded by (a function linear in) $\mathrm{VCD}(\mathcal{C})$.

Some connections between sample compression and teaching have been established in the literature (Doliwa et al., 2014; Darnstädt et al., 2016). The non-clashing property bears some similarities to sample compression and has in fact been used in the context of *unlabelled* sample compression (in which $f(S)$ is an unlabelled set) (Kuzmin and Warmuth, 2007; Chalopin et al., 2022). It is thus

natural to ask whether CN is an immediate upper or lower bound on NCTD. Using results from the literature, this question is answered negatively in a straightforward manner.

Proposition 40

1. *For every $k \in \mathbb{N}$, $k \geq 1$, there is a concept class $\mathcal{C}$ such that $\mathrm{NCTD}(\mathcal{C}) = \mathrm{PBTD}(\mathcal{C}) = 1$ but $\mathrm{CN}(\mathcal{C}) > k$.*

2. *Let $\mathcal{P}_m$ be the powerset over a domain of size m, where $m \geq 5$ is odd. Then $\mathrm{CN}(\mathcal{P}_m) < \mathrm{NCTD}(\mathcal{P}_m)$ and $2\mathrm{CN}(\mathcal{P}_m) < \mathrm{PBTD}(\mathcal{P}_m)$.*

Proof Statement 1 is due to the existence of a concept class $\mathcal{C}$ with $\mathrm{NCTD}(\mathcal{C}) = \mathrm{DTD}(\mathcal{C}) = 1$ and $\mathrm{VCD}(\mathcal{C}) = 5k$, see also Goldman and Kearns (1995). The fact that $\mathrm{CN}(\mathcal{C}) > k$ follows from a result by Floyd and Warmuth (1995) that states that no concept class of VC-dimension d has a sample compression scheme of size at most $\frac{d}{5}$.

Statement 2 follows from the obvious fact that $\mathrm{PBTD}(\mathcal{P}_m) = m$, in combination with Theorem 29, as well as with a result[4] by Darnstädt et al. (2016) that shows $\mathrm{CN}(\mathcal{P}_m) \leq \lfloor \frac{m}{2} \rfloor$, for any $m \geq 4$. ∎

Note that the compression function f in a sample compression scheme for $\mathcal{C}$ trivially induces a teacher mapping T_f defined by $T_f(C) = f(\{(x, C(x)) \mid x \in X\})$. The decompression mapping g then satisfies $g(T_f(C)) = C$ for all $C \in \mathcal{C}$. Hence (T_f, g) is a successful teacher-learner pair. Proposition 40.2 now states that there are concept classes for which the teacher-learner pairs (T_f, g) induced by any optimal sample compression scheme necessarily display collusion. In other words, optimal sample compression yields collusive teaching. An interesting problem is to find more examples of concept classes for which optimal sample compression yields collusive teachers and to determine necessary or sufficient conditions on the structure of such classes. Moreover, at present we do not know how large the gap between sample compression scheme size and NCTD can be.

As mentioned above, representation maps, which were proposed by Kuzmin and Warmuth (2007) and Chalopin et al. (2022), yield non-clashing teacher mappings. Clearly, in an unlabelled compression scheme, the representation map that compresses any concept in a class $\mathcal{C}$ to a subset of $\mathcal{X}$ must be injective, so that any two concepts in $\mathcal{C}$ remain distinguishable after compression. In other words, the non-clashing teacher mappings induced by representation maps are *repetition-free*, i.e., they do not map any two distinct concepts $C, C' \in \mathcal{C}$ to labelled samples $T(C), T(C')$ for which

$$\{x \in \mathcal{X} \mid (x, l) \in T(C) \text{ for some } l \in \{0, 1\}\} \neq \{x \in \mathcal{X} \mid (x, l') \in T(C') \text{ for some } l' \in \{0, 1\}\}.$$

Requiring no-clash teacher mappings to be repetition-free would be a limitation, as the example of the powerset over any set of m instances, $m \geq 2$, shows. In this case, no-clash teaching can be done with teacher mappings of order $\lceil \frac{m}{2} \rceil$, but it is not hard to see that the best possible repetition-free no-clash teacher mapping is of order m.

---

4. When $m = 5k$ for some $k \geq 1$, Darnstädt et al. (2016) even show that $\mathrm{CN}(\mathcal{P}_m) \leq 2k$; hence there is a family of concept classes with $\mathrm{CN} < \mathrm{NCTD}$ for which the gap between CN and NCTD grows linearly with the size of the instance space.

5. Strongly Collusion-Avoiding Teaching

Non-clashing teaching, while satisfying weak collusion-avoidance, displays a property that might be undesirable in practical settings, namely that a concept class may have two distinct non-clashing teacher mappings with exactly the same range. For example, consider the concept class $\mathcal{C}_G = \{\{i, i+1\} \mid i \in \mathbb{Z}\}$ and the teacher mappings T_1 and T_2 defined by $T_1(\{i, i+1\}) = \{(i, 1)\}$ and $T_2(\{i, i+1\}) = \{(i+1, 1)\}$. Clearly both mappings are non-clashing and of order 1. However, the range of T_1 is equal to the range of T_2, and yet T_1 and T_2 are distinct; in fact there is not even a single concept that these two teachers teach the same way. A learner, to be successful, will have to disambiguate between these two teachers, which could intuitively be considered some form of collusion.

For infinite concept classes, even preference-based teaching admits this kind of ambiguity. Consider for example the concept classes $\mathcal{C}_F = \{\{i, i+1\} \mid i \in \mathbb{N}\}$ and $\mathcal{C}_G = \{\{i, i+1\} \mid i \in \mathbb{Z}\}$ which were already introduced above. There are multiple preference-based teacher mappings realizing $\mathrm{PBTD}(\mathcal{C}_F) = 1$ and $\mathrm{PBTD}(\mathcal{C}_G) = 1$—in the latter case with identical ranges.

We are motivated to consider a strengthened notion of collusion avoidance that rules out this kind of ambiguity by insisting that teacher mappings have a unique inverse (restricted to their range). More precisely,

Definition 41 (Strong Collusion-Avoidance) *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$ and T a teacher mapping for $\mathcal{C}$. The teacher mapping T is strongly collusion-avoiding on a subset $\mathcal{C}'$ of $\mathcal{C}$, if there does not exist a different teacher mapping T' for $\mathcal{C}$ with $T(\mathcal{C}') \subseteq T'(\mathcal{C}')$.*

Remark 42 Note that (i) the concept class $\{[x, x+1) \mid x \in \mathbb{Q}\}$ admits a teacher mapping of order 1 that is strongly collusion-avoiding on all finite subsets, but no finite order teacher mapping that is strongly collusion-avoiding on the full class; and more generally (ii) if there exists a subset $\mathcal{C}' \subseteq \mathcal{C}$ such that no concept in $\mathcal{C}'$ has a definitive teaching set (with respect to $\mathcal{C}'$) of size k, then there is no teacher mapping of order k that is strongly collusion-avoiding on $\mathcal{C}'$.

We now state the main result of this section.

Theorem 43 *Let $\mathcal{C}$ be any concept class. There is a teacher mapping of order k that is strongly collusion-avoiding on all finite subsets of $\mathcal{C}$ if and only if $\mathrm{PBTD}(\mathcal{C}) \leq k$.*

Proof Consider the bipartite graph $B_{\mathcal{C}, \mathcal{X}}^{(k)}$, with elements of $\mathcal{C}$ (black vertices) and $\mathcal{C}$-samples of size k over $\mathcal{X}$ (white vertices), and an edge from $C_i \in \mathcal{C}$ to S_j whenever C_i is consistent with S_j. (Recall that any normal-form teacher mapping T of order k can be viewed as a matching M_T in $B_{\mathcal{C}, \mathcal{X}}^{(k)}$ that saturates all of the black vertices and a subset of the white vertices.)

Suppose that the teacher mapping T of order k is strongly collusion-avoiding on all finite subsets of $\mathcal{C}$. Let $\succ_T$ denote the relation on $\mathcal{C}$ in which $C \succ_T C'$ if there is an alternating path in $B_{\mathcal{C}, \mathcal{X}}^{(k)}$, from the vertex corresponding to C to the vertex corresponding to C', starting with an edge of M. Then $\succ_T$ is a strict order on $\mathcal{C}$, since the vertices of $\mathcal{C}$ belonging to any alternating cycle in $B_{\mathcal{C}, \mathcal{X}}^{(k)}$ would correspond to a finite subset of of $\mathcal{C}$ on which T is not strongly collusion-avoiding. Hence, $T(\mathcal{C})$ provides a $\succ_T$-distinguished teaching set for $\mathcal{C}$, and so $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ_T) = \mathcal{C}$, which implies $\mathrm{PBTD}(\mathcal{C}) \leq k$.

On the other hand, suppose that $\mathrm{PBTD}(\mathcal{C}) \leq k$, i.e., $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ) = \mathcal{C}$ for some strict order $\succ$, and let T be a teacher mapping that assigns a $\succ$-distinguished teaching set to each element of

$\mathcal{C}$. Let $\mathcal{C}'$ be any finite subset of $\mathcal{C}$ and suppose that T' is another teacher mapping on $\mathcal{C}$ that has $T'(\mathcal{C}') = T(\mathcal{C}')$. Then there must be a cycle in $B^{(k)}_{\mathcal{C},\mathcal{X}}$, restricted to vertices associated with $\mathcal{C}'$ and $T(\mathcal{C}')$, that alternates edges in the matching M associated with T and the matching M' associated with T'. But since any vertex C' reached by an alternating path starting at a vertex C in $\mathcal{C}'$ using an edge of M must satisfy $C \succ C'$ (by the transitivity of $\succ$), a contradiction. Hence, no such mapping T' exists, and thus T is a strongly collusion-avoiding teacher mapping of order k on all finite subsets of $\mathcal{C}$. ∎

An immediate corollary is that, for finite concept classes, recursive teaching realizes the optimal complexity of strongly-collusion avoiding teaching:

Corollary 44 *Let $\mathcal{C}$ be any finite concept class. There is a teacher mapping of order k that is strongly collusion-avoiding on all of $\mathcal{C}$ if and only if $\mathrm{RTD}(\mathcal{C}) \leq k$.*

Recall that the concept class $\mathcal{C}_{F+} = \mathcal{C}_F \cup \{\emptyset\}$ has $\mathrm{RTD}(\mathcal{C}_{F+}) = \infty$. Nevertheless, the teacher mapping T of order 1 in which $T(\{i, i+1\}) = (i, 1)$ and $T(\emptyset) = \emptyset$ is strongly collusion-avoiding on *all of* $\mathcal{C}_{F+}$. It follows that Corollary 44 does not extend to all infinite concept classes.

Nevertheless, an argument similar to that used in the proof of Theorem 43 shows that there is a teacher mapping of order k that is strongly collusion-avoiding on *all of* $\mathcal{C}$ if and only if there is a *well-founded* strict order $\succ$ on $\mathcal{C}$ such that $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ) = \mathcal{C}$.

Theorem 45 *Let $\mathcal{C}$ be any concept class. There is a teacher mapping of order k that is strongly collusion-avoiding on all of $\mathcal{C}$ if and only if there is a well-founded strict order $\succ$ on $\mathcal{C}$ such that $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ) = \mathcal{C}$.*

Proof Consider again the bipartite graph $B^{(k)}_{\mathcal{C},\mathcal{X}}$. Suppose that the teacher mapping T of order k is strongly collusion-avoiding on all of $\mathcal{C}$. Let $\succ_T$ denote the relation on $\mathcal{C}$ in which $C \succ_T C'$ if there is an alternating path in $B^{(k)}_{\mathcal{C},\mathcal{X}}$, from the vertex corresponding to C to the vertex corresponding to C', starting with an edge of the matching M_T. Since the vertices of $\mathcal{C}$ belonging to any alternating cycle in $B^{(k)}_{\mathcal{C},\mathcal{X}}$ would correspond to a finite subset of of $\mathcal{C}$ on which T is not strongly collusion-avoiding (since replacing the matched edges on such a cycle with the unmatched edges would give rise to a new teacher mapping with the same range as T), it follows that $\succ_T$ is a strict order on $\mathcal{C}$ with $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ) = \mathcal{C}$. Furthermore, let $C \in \mathcal{C}$ be any concept, and suppose there is an infinite alternating path P from C in $B^{(k)}_{\mathcal{C},\mathcal{X}}$ starting with the edge of M_T that saturates C. If there is an infinite alternating path from C in $B^{(k)}_{\mathcal{C},\mathcal{X}}$ starting with an unmatched edge, let P' be any such path. Alternatively, let P' be any alternating path from C in $B^{(k)}_{\mathcal{C},\mathcal{X}}$ starting with an unmatched edge and ending at an unsaturated white vertex. In either case, replacing matched edges with unmatched edges in the alternating path formed by concatenating P and P' would give rise to a new teacher mapping T' whose range contains the range of T. Since T is strongly collusion-avoiding on all of $\mathcal{C}$, there can be no such path P, and so $\succ_T$ is also well-founded.

On the other hand, suppose that $\mathrm{PBT}^{(k)}(\mathcal{C}, \succ) = \mathcal{C}$ for some well-founded strict order $\succ$, and let T be a teacher mapping of order k that assigns a $\succ$-distinguished teaching set to each element of $\mathcal{C}$. Suppose that T' is another teacher mapping of order k on $\mathcal{C}$ that has $T'(\mathcal{C}) \subseteq T(\mathcal{C})$, and let $\mathcal{C}' = \{C \in \mathcal{C} \mid T'(C) \neq T(C)\}$.

Let $C \in \mathcal{C}'$ and consider paths in $B_{\mathcal{C},\mathcal{X}}^{(k)}$ starting at C and alternating edges from M_T and $M_{T'}$. Since $\succ$ is well-founded all such paths that start with an edge of M_T are finite in length. Thus any such path must contain an alternating cycle, contradicting the fact that $\succ$ is a strict order on $\mathcal{C}$.

Hence, no such teacher mapping T' exists, and thus T is strongly collusion-avoiding on all of $\mathcal{C}$.

∎

6. Unifying Characterizations of Collusion-Avoiding Teaching Notions

6.1 Constrained Preference-Based Teaching and Matchings

The frameworks of preference-based teaching and bipartite matching provide natural and complementary bases for understanding the differences between the various forms of batch teaching discussed in earlier sections.

6.1.1 CONSTRAINED PREFERENCES

Let $\succ$ be any relation on $\mathcal{C}$. For any $C \in \mathcal{C}$, denote by $\succ$-preferred$_k(C)$ the set of C-samples of size k that form $\succ$-distinguished teaching sets for C. The relation $\succ$ is an *order k teacher-basis* for $\mathcal{C}$ if $\succ$-preferred$_k(C) \neq \emptyset$, for all $C \in \mathcal{C}$. Associated with any order k teacher-basis $\succ$ for $\mathcal{C}$ is a family of order k teacher mappings in which each concept C is mapped to an element of $\succ$-preferred$_k(C)$.

Suppose that $\succ$ is an order k teacher basis for $\mathcal{C}$ and $T_\succ$ is any teacher mapping associated with $\succ$. Then
(i) if $\succ$ is asymmetric, then $T_\succ$ is non-clashing (equivalently, weakly collusion-free) teacher mapping, so $\mathrm{NCTD}(\mathcal{C}) \leq k$ (cf. Theorem 13);
(ii) if $\succ$ is acyclic, then $T_\succ$ is a preference-based teacher mapping, so $\mathrm{PBTD}(\mathcal{C}) \leq k$;
(iii) if $\succ$ is well-founded, then $T_\succ$ is a strongly collusion-free teacher mapping (cf. Theorem 45);
(iv) if $\succ$ is strongly well-founded, then $T_\succ$ is a recursive teacher mapping, so $\mathrm{RTD}(\mathcal{C}) \leq k$ (cf. Proposition 11); and
(v) if $\succ$ is empty, then $T_\succ$ is a definitive teacher mapping.

Note that for finite concept classes, implications (ii), (iii), and (iv) collapse to:
(vi) if $\succ$ is acyclic, then $T_\succ$ is a recursive teacher mapping.

In the context of these characterizations, it should be noted that our paper is not the first to express teaching models by means of preference relations. Mansouri et al. (2019) formulated a framework of sequential teaching with preference relations. Depending on the constraints on the preference relation, this framework permits the re-expression of various notions of batch teaching studied in our paper, at least for concept classes over finite domains.

6.1.2 CONSTRAINED MATCHINGS

Consider once more the bipartite graph $B_{\mathcal{C},\mathcal{X}}^{(k)}$, with elements of $\mathcal{C}$ (black vertices) and $\mathcal{C}$-samples of size k over $\mathcal{X}$ (white vertices), and an edge from $C_i \in \mathcal{C}$ to S_j whenever C_i is consistent with S_j. Recall that if T is any order k teacher mapping (in normal form) for $\mathcal{C}$, then there is an associated full matching M_T in $\mathcal{B}_{\mathcal{C},\mathcal{X}}^{(k)}$. Any full matching M in $\mathcal{B}_{\mathcal{C},cX}^{(k)}$ induces a transitive relation $\succ_M$ on $\mathcal{C}$, where $C \succ C'$ if there is an alternating path with respect to M in $\mathcal{B}_{\mathcal{C},cX}^{(k)}$, from C to C', starting with an edge of M.

Suppose that T is any order k teacher mapping (in normal form) for $\mathcal{C}$. Then

(i) if $\mathrm{NCTD}(\mathcal{C}) \leq k$ is certified by the non-clashing mapping T, then $\succ_{M_T}$ is asymmetric (equivalently, $\mathcal{B}_{\mathcal{C},cX}^{(k)}$ contains no alternating cycle of length 4 with respect to M_T);

(ii) if $\mathrm{PBTD}(\mathcal{C}) \leq k$ is certified by the preference-based teacher mapping T, then $\succ_{M_T}$ is a strict order on $\mathcal{C}$ (equivalently, $\mathcal{B}_{\mathcal{C},cX}^{(k)}$ contains no alternating cycle with respect to M_T, i.e. M_T is a uniquely restricted matching (Golumbic et al., 2001); cf. Theorem 43);

(iii) if T is strongly collusion-free, then $\succ_{M_T}$ is a well-founded strict-order on $\mathcal{C}$ (equivalently, $\mathcal{B}_{\mathcal{C},cX}^{(k)}$ contains no unbounded alternating path with respect to M_T);

(iv) if $\mathrm{RTD}(\mathcal{C}) \leq k$ is certified by the recursive teacher mapping T, then $\succ_{M_T}$ is a strongly well-founded strict-order on $\mathcal{C}$ (equivalently, every alternating path with respect to M_T in $\mathcal{B}_{\mathcal{C},cX}^{(k)}$, starting from a concept C, is bounded in length by some constant (depending on C); and

(v) if if $\mathrm{RTD}(\mathcal{C}) \leq k$ is certified by the recursive teacher mapping T, then $\succ_{M_T}$ is empty (equivalently, the matching M_T is an induced matching in $\mathcal{B}_{\mathcal{C},cX}^{(k)}$).

Note that for finite concept classes, implications (ii), (iii), and (iv) collapse to:

(vi) if $\mathrm{RTD}(\mathcal{C}) \leq k$ is certified by the recursive teacher mapping T, then $\succ_{M_T}$ is a strict order on $\mathcal{C}$ (equivalently, $\mathcal{B}_{\mathcal{C},cX}^{(k)}$ contains no alternating cycle with respect to M_T.

These characterizations in terms of constrained bipartite matching problems have implications on the computational complexity of decision problems relating to teaching complexity parameters; the interested reader is referred to (Kirkpatrick et al., 2019) for more details.

6.2 Broadened Recursion

Recall, from Subsection 2.2, that the notion of Recursive Teaching, which involves the iterative removal of maximal Definitive Teaching sets, has a natural generalization involving the iterative removal of Recursive Teaching sets.

In Recursive Teaching, in the first iteration, one removes all concepts with smallest definitive teaching dimension from the concept class. In the next iteration, one removes all concepts with smallest definitive teaching dimension from the remainder of the concept class. This process is repeated for as long as there are concepts remaining.

We can now view Recursive Teaching as the first layer of a hierarchy of teaching notions. In the $(d+1)$st layer, one again proceeds iteratively, always removing all concepts with smallest "dth layer recursive teaching dimension".

We make this notion of Doubly-Recursive Teaching more precise in the following:

Definition 46 *Let $\mathcal{C}$ be a concept class and let $2\text{-}\mathrm{RT}_{0,*}^{(k)}(\mathcal{C}) = \mathrm{DT}^{(k)}(\mathcal{C})$. For all $i, d \in \mathbb{N}$ define*

$$2\text{-}\mathrm{RT}_{d+1,0}^{(k)}(\mathcal{C}) = 2\text{-}\mathrm{RT}_{d,*}^{(k)}(\mathcal{C})$$

and

$$2\text{-}\mathrm{RT}_{d+1,i+1}^{(k)}(\mathcal{C}) = 2\text{-}\mathrm{RT}_{d+1,i}^{(k)}(\mathcal{C}) \cup 2\text{-}\mathrm{RT}_{d,*}^{(k)}(\mathcal{C} \setminus 2\text{-}\mathrm{RT}_{d+1,i}^{(k)}(\mathcal{C}))$$

where $2\text{-}\mathrm{RT}_{d+1,*}^{(k)}(\mathcal{C}) = \cup_{i \geq 0} 2\text{-}\mathrm{RT}_{d+1,i}^{(k)}(\mathcal{C})$.

Note that $2\text{-}\mathrm{RT}_{0,*}^{(k)}(\mathcal{C}) = \mathrm{DT}^{(k)}(\mathcal{C})$ and $2\text{-}\mathrm{RT}_{1,i}^{(k)}(\mathcal{C}) = \mathrm{RT}_i^{(k)}(\mathcal{C})$ (and hence $2\text{-}\mathrm{RT}_{1,*}^{(k)}(\mathcal{C}) = \mathrm{RT}_*^{(k)}(\mathcal{C})$), for all $\mathcal{C}$. Furthermore, $2\text{-}\mathrm{RT}_{d,*}^{(k)}(\mathcal{C}) = \emptyset$ if and only if $\mathrm{DT}^{(k)}(\mathcal{C}) = \emptyset$. It is straightfor-

ward to confirm that 2-RT$_{2,*}^{(1)}(\mathcal{C}_{F+}) = \mathcal{C}_{F+}$, that is $\mathcal{C}_{F+}$ has a teaching mapping of order 1, realized by a doubly-recursive scheme of depth 2. More generally, the same is true of the concept class $\mathcal{C}_{H2}$ which consists of infinitely many infinite size concepts (over the ground set $2^{\mathbb{N}x\mathbb{N}}$) given by $\mathcal{C}_{H2} = \{S_{i,j} \mid i, j \in \mathbb{N}\}$, where $S_{i,j} = \{\{(j, i), (j + 1, i)\} \cup \mathbb{N} \times \{i + 1\}\}$. The reader should have no difficulty generalizing the concept class $\mathcal{C}_{H2}$ to a class $\mathcal{C}_{Hd}$ (over the ground set $2^{\mathbb{N}^d}$) and to note that $\mathcal{C}_{Hd}$ has a teaching mapping of order 1, realized by a doubly-recursive scheme of depth d, but no finite order teacher mapping using doubly-recursive schemes of depth less than d.

Observe that the teacher mapping implicit in the definition of 2-RT$_{d,*}^{(k)}(\mathcal{C})$ is strongly collusion-avoiding. This is an immediate consequence of the fact that every 2-recursive teaching plan is made up of a succession of maximal size recursive teaching sets, which themselves are made up of a succession of definitive teaching sets. However, there exist concept classes $\mathcal{C}$ with strongly collusion-avoiding teacher mappings of order k for which 2-RT$_{d,*}^{(k)}(\mathcal{C}) \neq \mathcal{C}$, for all $d \geq 0$.

7. Conclusions

No-clash teaching represents the limit of data efficiency that can be achieved in batch teaching settings obeying Goldman and Mathias's criterion for collusion-avoidance. Therefore, it is the sole most promising such teaching model to shed light on two open problems in computational learning theory, namely (i) to find a teaching complexity parameter that is upper-bounded by a function linear in VCD, and (ii) to help establish new results towards an upper bound on the size of smallest sample compression schemes that is linear in VCD. If *any* teaching model that satisfies the weak collusion-avoiding criterion yields a complexity upper-bounded by (a function linear in) VCD, then no-clash teaching does. Likewise, if *any* such teaching model is powerful enough to compress concepts as effectively as sample compression schemes do, then no-clash teaching is. The most fundamental open question resulting from our paper is probably whether NCTD is upper-bounded by VCD in general.

An interesting related question is how NCTD relates to PBTD and RTD. With the parity functions, we have found a family of concept classes for which PBTD and RTD exceed NCTD by a factor of four, but so far we have not found a family of concept classes for which PBTD and RTD exceed any function linear in NCTD. Whether or not such a family exists is another open problem.

Our paper also raises the question how various notions of collusion-avoidance constrain the interaction between teacher and learner and thus affect the teaching complexity. We have demonstrated that weak collusion-avoidance is a defining property of non-clashing teaching, as is strong collusion-avoidance for preference-based teaching. Likewise, other notions of collusion-avoidance may give rise to new teaching models or characterize existing teaching complexity notions. In this sense, our results might be a stepping stone for a broader study on collusion-avoidance in teaching.

Acknowledgments

the Canada CIFAR AI Chairs program, as an Affiliate Chair with the Alberta Machine Intelligence Institute (Amii).

Appendix A. Examples of Concept Classes Satisfying $\mathrm{NCTD} = 1$ and $\mathrm{PBTD} = 3$

Example 2 *The concept class $\mathcal{C}_{3/1}$ shown in Table 3 satisfies $\mathrm{NCTD}^+(\mathcal{C}_{3/1}) = 1$, as witnessed by the entries highlighted in bold corresponding to non-clashing teaching sets. Consequently, one also obtains $\mathrm{NCTD}(\mathcal{C}_{3/1}) = 1$. By contrast, we have $\mathrm{PBTD}(\mathcal{C}_{3/1}) \geq 3$, which can be verified by observing that every labeled pair of instances occurs in two concepts in $\mathcal{C}_{3/1}$.*

When restricting this concept class to the instances x_1 through x_4, as shown in Table 4, the RTD (and thus PBTD) remains 3. While NCTD^+ increases to 2, one still obtains an NCTD of 1, again witnessed by the highlighted entries.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
C_1	**1**	1	0	1	0	1	0	0
C_2	0	**1**	0	0	0	1	1	1
C_3	1	1	**1**	0	0	0	0	1
C_4	0	1	1	**1**	0	0	1	0
C_5	1	0	1	1	**1**	0	0	0
C_6	0	0	0	1	1	**1**	1	0
C_7	0	0	1	0	1	0	**1**	1
C_8	1	0	0	0	1	1	0	**1**

Table 3: The concept class $\mathcal{C}_{3/1}$ with $\mathrm{NCTD}^+(\mathcal{C}_{3/1}) = 1$ and $\mathrm{PBTD}(\mathcal{C}_{3/1}) = 3$.

	x_1	x_2	x_3	x_4
C_1	**1**	1	0	1
C_2	0	**1**	0	0
C_3	1	1	**1**	0
C_4	0	1	1	**1**
C_5	1	**0**	1	1
C_6	0	0	**0**	1
C_7	**0**	0	1	0
C_8	1	0	0	**0**

Table 4: The concept class $\mathcal{C}_{3/1}$ restricted to the first four instances, which results in an NCTD of 1 and a PBTD of 3.

One can argue that any finite concept class with $\mathrm{NCTD} = 1$ and $\mathrm{PBTD} = 3$ must have at least eight concepts and four instances: If a concept class has seven or fewer concepts, then its VC-dimension is at most 2, and a simple argument will show that its PBTD is at most 2 as well.[5] If it has only three instances, yet eight concepts, it is the powerset and thus has an NCTD of 2.

---

5. The smallest concept class for which PBTD exceeds VCD has 10 concepts (Doliwa et al., 2014).

Thus, $\mathcal{C}_{3/1}$ restricted to the first four instances is a smallest concept class with NCTD $= 1$ and PBTD $= 3$.

An additional example of a concept class with NCTD $= 1$ and PBTD $= 3$ that is interesting due to its connection with the theory of combinatorial designs, is the so-called Kummer configuration.

Example 3 *The Kummer configuration is a configuration of 16 points and 16 planes such that*
(i) each point lies on exactly 6 planes,
(ii) each plane contains exactly 6 of the points,
(iii) each pair of planes intersects in exactly 2 of the points, and
(iv) and each pair of points is contained in exactly two of the planes.

As described by Assmus and Sardi 1981, this configuration can be interpreted as a concept class $\mathcal{C}_K$ of 16 concepts over 16 instances. Each Kummer instance (point) can be thought of as a pair (i, j) for $0 \leq i, j \leq 3$, while each Kummer concept (plane) $C_{a,b}$ for $0 \leq a, b \leq 3$ contains the three instances (a, j) for $j \neq b$ as well as the three instances (i, b) for $i \neq a$. (See Figure 1 (left) for an illustration of the Kummer concept $C_{1,1}$.)
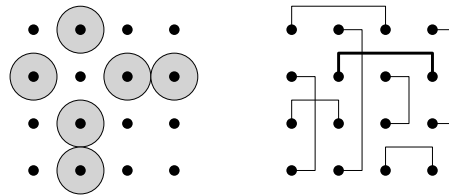

Figure 1: The Kummer concept $C_{1,1}$ (left) and teacher mapping (right)

It is easy to see that each Kummer concept can be uniquely determined by three labelled instances: for example, $C_{i,j}$ is uniquely determined by the set $\{((i, j), -), ((i + 1) \mod 4, j), +), ((i, (j + 1) \mod 4), +)\}$. *Furthermore, no two labelled instances suffice, since (i) every pair of positively labelled instances is consistent with two concepts, and (ii) every other pair of labelled instances is consistent with zero or (at least) two concepts. Thus,* $\text{PBTD}(\mathcal{C}_K) = \text{DTD}(\mathcal{C}_K) = 3$.

To see that $\text{NCTD}^+(\mathcal{C}_K) = 1$ we define a symmetric matching between concepts and instances (if $T((C_{i,j}) = (i', j')$ then $T((C_{i',j'}) = (i, j))$. One such matching is represented in Figure 1 (right), with the edge corresponding to the symmetric pair $T((C_{1,1}) = (1, 3)$ and $T((C_{1,3}) = (1, 1)$ highlighted in bold. That this mapping is non-clashing follows as an immediate consequence of the fact that no two edges span the same rows or columns.

References

Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

Brenna Argall, Sonia Chernova, Manuela M. Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.

Edward F. Assmus and J.E.Ñovillo Sardi. Generalized steiner systems of type 3-(v,4,6,1). In *Proceedings of the Finite Geometries and Designs Conference*, pages 16–21. Cambridge University Press, 1981.

Frank Balbach. Measuring teachability using variants of the teaching dimension. *Theoretical Computer Science*, 397(1–3):94–113, 2008.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, pages 41–48, 2009.

Jérémie Chalopin, Victor Chepoi, Shay Moran, and Manfred K. Warmuth. Unlabeled sample compression schemes and corner peelings for ample and maximum classes. *Journal of Computer and System Sciences*, 127:1–28, 2022.

Malte Darnstädt, Thorsten Kiss, Hans Ulrich Simon, and Sandra Zilles. Order compression schemes. *Theoretical Computer Science*, 620:73–90, 2016.

Colin de la Higuera. Characteristic sets for polynomial grammatical inference. *Machine Learning*, 27(2):125–138, 1997.

François Denis. Learning regular languages from simple positive examples. *Machine Learning*, 44 (1/2):37–66, 2001.

Thorsten Doliwa, Gaojian Fan, Hans Ulrich Simon, and Sandra Zilles. Recursive teaching dimension, VC-dimension and sample compression. *Journal of Machine Learning Research*, 15: 3107–3131, 2014.

Sally Floyd and Manfred Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(3):1–36, 1995.

Ziyuan Gao, Christoph Ries, Hans Ulrich Simon, and Sandra Zilles. Preference-based teaching. *Journal of Machine Learning Research*, 18(31):1–32, 2017.

Sally A. Goldman and Michael J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995.

Sally A. Goldman and H. David Mathias. Teaching a smarter learner. *Journal of Computer and System Sciences*, 52(2):255–267, 1996.

Martin C. Golumbic, Tirza Hirst, and Moshe Lewenstein. Uniquely restricted matchings. *Algorithmica*, 31(2):139–154, 2001.

David Haussler, Nick Littlestone, and Manfred K. Warmuth. Predicting $\{0, 1\}$ functions on randomly drawn points. *Information and Computation*, 115(2):284–293, 1994.

Mark K Ho, Michael Littman, James MacGlashan, Fiery Cushman, and Joseph L Austerweil. Showing versus doing: Teaching by demonstration. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 3027–3035. 2016.

Lunjia Hu, Ruihan Wu, Tianhong Li, and Liwei Wang. Quadratic upper bound for recursive teaching dimension of finite VC classes. In *Proceedings of the 30th Conference on Learning Theory (COLT)*, pages 1147–1156, 2017.

David Kirkpatrick, Hans U. Simon, and Sandra Zilles. Optimal collusion-free teaching. In *Proceedings of the 34th International Conference on Algorithmic Learning Theory (ALT)*, volume 98, pages 506–528, 2019.

Dima Kuzmin and Manfred K. Warmuth. Unlabeled compression schemes for maximum classes. *Journal of Machine Learning Research*, 8:2047–2081, 2007.

Steffen Lange, Thomas Zeugmann, and Sandra Zilles. Learning indexed families of recursive languages from positive data: A survey. *Theoretical Computer Science*, 397(1-3):194–232, 2008.

Nick Littlestone and Manfred K. Warmuth. Relating data compression and learnability. Technical Report, UC Santa Cruz, 1986.

Farnam Mansouri, Yuxin Chen, Ara Vartanian, Jerry Zhu, and Adish Singla. Preference-based batch and sequential teaching: Towards a unified view of models. In *Advances in Neural Information Processing Systems (NeurIPS) 32*, pages 9199–9209, 2019.

Shay Moran and Amir Yehudayoff. Sample compression schemes for VC classes. *Journal of the ACM*, 63(3):21:1–21:10, 2016.

Shay Moran, Amir Shpilka, Avi Wigderson, and Amir Yehudayoff. Teaching and compressing for low VC-dimension. *CoRR*, abs/1502.06187, 2015.

Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13 (1):145–147, 1972.

Ingo Schwab, Wolfgang Pohl, and Ivan Koychev. Learning to recommend from positive evidence. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI)*, pages 241–247, 2000.

Patrick Shafto, Noah D. Goodman, and Thomas L. Griffiths. A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71:55–89, 2014.

Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Generation Computing*, 8:337–348, 1991.

Hans Ulrich Simon. Minimum tournaments with the strong s_k-property and implications for teaching. *CoRR*, abs/2205.08357, 2022.

Vladimir N. Vapnik and Alexey Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.

Chunlin Wang, Chris H. Q. Ding, Richard F. Meraz, and Stephen R. Holbrook. Psol: a positive sample only learning algorithm for finding non-coding RNA genes. *Bioinformatics*, 22(21):2590–2596, 2006.

Kenneth Wexler and Peter W. Culicover. *Formal Principles of Language Acquisition*. MIT Press, 1980.

Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N. Rafferty. An overview of machine teaching. *CoRR*, abs/1801.05927, 2018.

Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *Journal of Machine Learning Research*, 12:349–384, 2011.