Matching-based Parameters in Machine Teaching: a Unified Framework and Hierarchy

Hans U. Simon* Jan Arne Telle†

May 19, 2026

Abstract

We investigate structural relationships between recently introduced matching-based parameters in Machine Teaching, focusing on the Saturating Matching Number (SMN), Greedy Matching Number (GMN), and Antichain Matching Number (AMN). The connection to matching arises from viewing Teaching Dimension as an injective mapping from concepts to small labeled sample sets. We present a unified framework that links these teaching complexity parameters with other combinatorial parameters, enabling a systematic comparison with classical measures such as VC-dimension (VCD), Recursive Teaching Dimension (RTD), and Subset Teaching Dimension for minimum sets ($\mathrm{STD}_{\mathrm{min}}$). Within this framework, we establish a hierarchy of inequalities among the parameters. We further analyze key structural properties, including class- and domain-monotonicity under alternative non-collusion teaching constraints, and characterize the behavior of the parameters on canonical concept classes such as the power-set. Finally, we study their behavior under class composition, identifying conditions for additivity, sub-additivity, and super-additivity. Our results contribute to a deeper understanding of the combinatorial foundations of machine teaching and its connections to classical learning-theoretic complexity measures.

1 Introduction

In formal models of *machine learning* [19] we have a binary concept class C of possible hypotheses, an unknown target concept $c \in C$ and training data in the form of correctly labeled randomly sampled examples that are given to the learner. The concept class contains classification rules (called concepts) which split the instances in its domain X into positive examples (labeled 1) and negative examples (labeled 0). In formal models of *machine teaching*, the correctly labeled example set S_c given to the learner is instead carefully chosen by a teacher. The set S_c is called a teaching set for the target concept c in the class C, and the

*Faculty of Mathematics, Ruhr-University Bochum, Germany. E-mail: hans.simon@rub.de.
†Department of Informatics, University of Bergen, Norway. E-mail: jan.arne.telle@uib.no.

goal is to find a small teaching set that allows the learner (or learner algorithm) to "infer" c. A common goal in machine teaching is to prove bounds on the *teaching dimension* of the class C, i.e. the smallest value k so that for every concept $c \in C$ there exists a teaching set S_c of cardinality at most k allowing to uniquely identify c.

In recent years, the field of machine teaching has seen various applications in fields like pedagogy [15], trustworthy AI [24], reinforcement learning [23], active learning [21] and explainable AI [22]. Various models of machine teaching have been proposed, e.g. the classical teaching dimension model [8], the optimal teacher model [1], recursive teaching [25], preference-based teaching [7], no-clash teaching [3], probabilistic teaching [5], with each of them giving rise to a new variation on teaching dimension. One major line of research has been to ask how parameters from teaching complexity relate to one another and to other interesting parameters from computational learning theory, like VC-dimension [20], the optimal mistake from online learning [11], or the size of sample compression schemes [6]. A big open problem in the field asks if the recursive teaching dimension RTD is always linear in the VC-dimension VCD, with the current best being a quadratic bound [9].

In this paper we continue this exploration, with a focus on three parameters related to matching problems that have recently appeared in the literature on machine teaching [4, 12, 17]. The connection to matching is as follows: to prove that the teaching dimension of a concept class C under a certain model of machine teaching has a value of k, we must provide an injective mapping from C to sets of at most k labeled examples, satisfying certain conditions depending on the model[1]. For a concept class C, by applying this connection to the sample complexity of MAP-based and MLE-based teaching introduced in a JMLR-article in 2024 [17] we achieve the first parameter that we call the Saturating Matching Number $\mathrm{SMN}(C)$. The second parameter, that we call the Greedy Matching Number $\mathrm{GMN}(C)$, was introduced in a ML-article in 2026 [4] as the largest sample complexity that may arise from Greedy teaching, while the third parameter, that we call the Antichain Matching Number $\mathrm{AMN}(C)$, was introduced in a NeurIPS-article in 2022 [12] as a lower bound on the sample complexity of any teaching map having the antichain property. In Section 2 we formally define all three parameters $\mathrm{SMN}, \mathrm{GMN}$ and AMN within a unified framework of matchings, to highlight their correspondences and simplify their comparison. We will also compare them to three fundamental combinatorial parameters SMN', GMN' and AMN' that bound their respective namesakes in an essential way, and also to the two well-known parameters VC-dimension VCD [20] and Recursive Teaching Dimension RTD [25]. Finally, we add a ninth parameter into the comparison, the Subset Teaching Dimension for minimum subsets STD_{min}, that was introduced in [12] as the first (batch) teaching parameter upper-bounded by VCD.

To avoid coding tricks between the teacher and the learner (e.g. agreeing on a linear ordering of concepts and examples and teaching the i'th concept using the i'th example) many teaching models have adhered to the Goldman-Mathias condition of collusion-freeness introduced in 1995[8]. Recently, some papers [4, 12] have argued that there are application settings in which this condition is not natural, and alternative conditions have been sug-

[1]Exact definitions will be given in the next section.

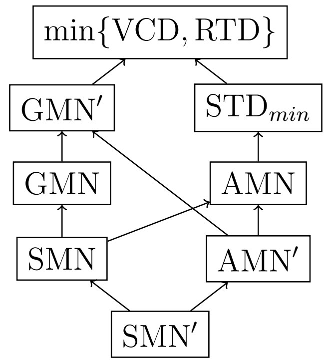

Figure 1: The parameter hierarchy: an arc $Y \to Z$ represents a relation $Y(C) \leq Z(C)$ that is valid for each concept class C and that is occasionally strict (for special choices of C).

gested. In particular, the STD_{min} model instead adheres to the three conditions of being class-monotonic, domain-monotonic, and of observing the antichain property [12].

The rest of our paper is organized as follows. In Section 2 we define the combinatorial parameters under consideration. In Section 3 we investigate several $\leq$-relations between these parameters, applying also results from a 2024 ICML-article [18] resolving an open problem in machine teaching related to SMN′. The resulting parameter hierarchy, called SMN-GMN-AMN hierarchy in what follows, is visualized in Fig. 1. In Section 4 we ask which of the parameters obey the conditions on teaching without coding tricks that were introduced in [12], and study their class-monotonicity and domain-monotonicity. In Section 5 we look at the behaviour of these parameters on the special case of the concept class being equal to the power-set, which is a classical focus in much of the literature. In Section 6, we determine which of our combinatorial parameters are additive (resp. sub- or super-additive) on the free combination of two concept classes. We show finally in Section 7 that all inequalities of the SMN-GMN-AMN hierarchy are occasionally strict so that this hierarchy is proper.

2 Preliminaries

We first recall a few definitions from learning theory. Let X be a finite set and let 2^X denote the set of all functions $c : X \to \{0,1\}$. A family $C \subseteq 2^X$ is called a *concept class over the domain* $X = \mathrm{dom}(C)$. A set $S = \{(x_1, b_1), \ldots, (x_m, b_m)\}$ with m distinct elements $(x_i, b_i) \in X \times \{0,1\}$ is called a *labeled sample of size* $|S| = m$. A concept $c \in C$ is said to be *consistent with* S if $c(x_i) = b_i$ for $i = 1, \ldots, m$. A labeled sample S is said to be *realizable by* C if C contains a concept that is consistent with S. A C*-saturating matching* is a mapping M that assigns to each $c \in C$ a labeled sample $M(c)$ such that the following holds: first c is consistent with $M(c)$; second $c \neq c' \in C$ implies that $M(c) \neq M(c')$. The *cost of* M is defined as the size of the largest labeled sample that is assigned by M to one of the concepts in C, i.e., $\mathrm{cost}(M) = \max_{c \in C} |M(c)|$. A C-saturating matching M' is called a *direct improvement* of another C-saturating matching M if M' differs from M only on a single concept c and $|M'(c)| < |M(c)|$. A C-saturating matching M is called *greedy* if it does not admit for direct improvements.

Suppose that P is a procedure that, given a concept class C, returns a C-saturating matching M. P is said to be *greedy* if it inspects the concepts one by one and, for each fixed concept $c \in C$, makes an assignment $M(c) = S$ where S is a labeled sample with the following properties:

1. The concept c is consistent with S.
2. The labeled sample S is still available, i.e., it has not already been assigned to a different concept.
3. Among the samples which meet the preceding two criteria, S is of lowest size.

A greedy procedure has some degrees of freedom. First, we have left open in which order the concepts are inspected. Second, for each concept, P may have the choice between different labeled samples of the same size. There are greedy procedures that have a unique output, for example procedures that commit to a linear order both among the concepts and among the labeled samples, where the latter order must be of non-decreasing size. Without such commitments, P may have several possible outputs, and for some machine teaching application this may in fact be desirable. Greedy C-saturating matchings and greedy procedures for their computation are related as follows:

Remark 2.1. *A C-saturating matching is greedy iff it is a possible output of a greedy procedure.*

This since for any greedy matching M we can order the concepts by the size of their assigned sample and run a greedy procedure on this order to get M as output, and moreover it follows by an easy induction that the output of a greedy procedure does not admit for direct improvements. In the sequel, the symbol X will always denote the domain of a concept class C. The latter will be clear from context. We are now prepared to define the parameters we are interested in.

SMN: The *saturating matching number of* C, denoted by $\mathrm{SMN}(C)$, is the cost of a cheapest C-saturating matching. This parameter was introduced first in [17]. It characterizes the sample complexity of MAP- resp. MLE-based teaching. See [17] for details.

SMN′: Here comes a lower bound on the saturating matching number. We define $\mathrm{SMN}'(C)$ as the smallest number d such that the number of C-realizable samples of size at most d is greater than or equal to $|C|$.

AMN: The *antichain matching number of* C, denoted by $\mathrm{AMN}(C)$, is the cost of a cheapest C-saturating matching that has the antichain property. The antichain matching number was introduced first in [12]. As explained in [12], it lower-bounds the sample complexity in all models where the teaching map has the antichain property (or can be modified so as to have this property without increasing its cost).

AMN′: We are also interested in the following lower bound on the antichain matching number. We define $\mathrm{AMN}'(C)$ as the smallest number d such that the C-realizable samples of size at most d contain an antichain of size $|C|$.

GMN: The parameter $\mathrm{GMN}(C)$, introduced first in [4] in connection with learning representations of concepts, is defined as the largest possible cost of a greedy C-saturating matching. It is easy to see that the smallest possible cost of a greedy C-saturating matching equals $\mathrm{SMN}(C)$.

GMN′: We define $\mathrm{GMN}'(C)$ as the smallest number d such that $\sum_{i=0}^{d} \binom{|X|}{i} \geq |C|$.

VCD: Points $x_1 \ldots, x_r \in X$ are said to be *shattered by* C if, for every choice of $b_1, \ldots, b_r \in \{0, 1\}$, the labeled sample $\{(x_1, b_1), \ldots, (x_r, b_r)\}$ is C-realizable. The size of the largest set that is shattered by C is called the *VC-dimension of* C and denoted by $\mathrm{VCD}(C)$. The VC-dimension was first introduced in [20].

RTD: By $\mathrm{RTD}(C)$, we denote the so-called *recursive teaching dimension of* C. We will not require a formal definition of the RTD in this paper, but the interested reader may consult [25, 2] for a definition.

STD$_{min}$: A sequence $\mathcal{M} = (M_k)_{k \geq 0}$ of C-saturating matchings is called a *subset teaching sequence for* C if the following hold:

1. $M_0(c) = \{(x, c(x)) : x \in X\}$ for all $c \in C$.
2. $M_{k+1}(c) \subseteq M_k(c)$ for all $k \geq 0$ and all $c \in C$.
3. $M_{k+1}(c) \not\subseteq M_k(c')$ for all $k \geq 0$ and all $c' \neq c \in C$.

Let
$$k^*(\mathcal{M}) = \min\{k : \forall k' \geq k, c \in C : M_{k'}(c) = M_k(c)\} \enspace .$$
We define $\mathrm{cost}(\mathcal{M})$ as the cost of the matching $M_{k^*(\mathcal{M})}$. Finally, we set
$$\mathrm{STD}_{min}(C) = \min\{\mathrm{cost}(\mathcal{M}) : \mathcal{M} \text{ is a subset teaching sequence for } C\} \enspace .$$
This parameter was introduced in [12]. It is a variant of Balbach's [1] subset-teaching dimension.

3 A First Look at the SMN-GMN-AMN Hierarchy

Here is a list of the (more or less) obvious relations among the parameters of the SMN-GMN-AMN hierarchy:

1. $\mathrm{SMN}(C) \leq \mathrm{GMN}(C)$ and $\mathrm{SMN}(C) \leq \mathrm{AMN}(C)$.
   Reason: No matching can be cheaper than a cheapest one.

2. $\mathrm{SMN}'(C) \le \mathrm{SMN}(C)$.
Reason: Let M be a C-saturating matching of cost $\mathrm{SMN}(C)$. Then the samples $M(c)$ with $c \in C$ are C-realizable. It follows that the number of C-realizable samples of size at most $\mathrm{SMN}(C)$ is greater than or equal to $|C|$. The definition of $\mathrm{SMN}'(C)$ now implies that $\mathrm{SMN}'(C) \le \mathrm{SMN}'(C)$.

3. $\mathrm{AMN}'(C) \le \mathrm{AMN}(C)$.
Reason: Let M be a C-saturating matching of cost $\mathrm{AMN}(C)$ that has the antichain property. Then the samples $M(c)$ with $c \in C$ are C-realizable and form an antichain of size $|C|$. The definition of $\mathrm{AMN}'(C)$ now implies that $\mathrm{AMN}'(C) \le \mathrm{AMN}(C)$.

4. $\mathrm{GMN}(C) \le \mathrm{GMN}'(C)$.
Reason: Set $d := \mathrm{GMN}'(C)$ and $X := \mathrm{dom}(C)$. It follows that $\sum_{i=0}^{d} \binom{|X|}{i} \ge |C|$. Let M be a C-saturating matching of cost $d' > d$. Pick a concept $c_0 \in C$ such that $|M(c_0)| = d'$. For cardinality reasons, there must exist an unlabeled sample $\{x_1, \ldots, x_r\}$ of size $r \le d$ such that no $c \in C$ satisfies $M(c) = \{(x_1, c(x_1)), \ldots, (x_r, c(x_r))\}$. Let M' be the matching that equals M except for $M'(c_0) = \{(x_1, c_0(x_1)), \ldots, (x_r, c_0(x_r))\}$. Then M' is a C-saturating matching that directly improves on M. We have therefore shown that no C-saturating matching of cost exceeding d is greedy. Hence $\mathrm{GMN}(C) \le \mathrm{GMN}'(C)$.

5. $\mathrm{SMN}'(C) \le \mathrm{AMN}'(C)$
Reason: If the C-realizable samples of size at most d contain an antichain of size $|C|$ then, obviously, the number of C-realizable samples of size at most d cannot be smaller than $|C|$. The inequality $\mathrm{SMN}'(C) \le \mathrm{AMN}'(C)$ is therefore evident from the definitions of $\mathrm{SMN}'(C)$ and of $\mathrm{AMN}'(C)$.

6. $\mathrm{AMN}(C) \le \mathrm{STD}_{min}(C)$ and $\mathrm{STD}_{min}(C) \le \min\{\mathrm{VCD}(C), \mathrm{RTD}(C)\}$.
Reason: It is obvious from the definition of a subset teaching sequence $\mathcal{M} = (M_k)_{k \ge 0}$ (and it was noted in [12] already) that each matching M_k is C-saturating and has the antichain property. Hence $\mathrm{AMN}(C) \le \mathrm{STD}_{min}(C)$. It was shown furthermore in [12] that $\mathrm{STD}_{min}(C)$ cannot exceed $\mathrm{VCD}(C)$ or $\mathrm{RTD}(C)$.

7. $\mathrm{GMN}'(C) \le \min\{\mathrm{VCD}(C), \mathrm{RTD}(C)\}$.
Reason: According to Sauer's Lemma [14, 16], we have that $\sum_{i=0}^{\mathrm{VCD}(C)} \binom{|X|}{i} \ge |C|$. This inequality holds as well with $\mathrm{RTD}(C)$ in place of $\mathrm{VCD}(C)$ [13]. The inequalities $\mathrm{GMN}'(C) \le \mathrm{VCD}(C)$ and $\mathrm{GMN}'(C) \le \mathrm{RTD}(C)$ are now immediate from the definition of $\mathrm{GMN}'(C)$.

8. If $\mathrm{SMN}'(C) \le |X|/5$, then $\mathrm{GMN}'(C) \le 2 \cdot \mathrm{SMN}'(C)$.
Reason: For sake of brevity, set $n = |X|$. Let d be the smallest number subject to $2^d \cdot \sum_{i=0}^{d} \binom{n}{i} \ge |C|$. Since the number of C-realizable samples of size at most d is bounded from above by $\sum_{i=0}^{d} 2^i \cdot \binom{n}{i} \le 2^d \cdot \sum_{i=0}^{d} \binom{n}{i}$, it follows from the definition of

$\mathrm{SMN}'(C)$ that $\mathrm{SMN}'(C) \geq d$. It suffices therefore to show $\mathrm{GMN}'(C) \leq 2d$. We will show that

$$\sum_{i=0}^{2d} \binom{n}{i} \geq 2^d \cdot \sum_{i=0}^{d} \binom{n}{i} \ . \tag{1}$$

Thanks to $2^d \cdot \sum_{i=0}^{d} \binom{n}{i} \geq |C|$, this would imply that $\mathrm{GMN}'(C) \leq 2d$. We verify (1) by proving that

$$\forall i = 0, \ldots, d : \frac{\binom{n}{d+i}}{\binom{n}{i}} \geq 2^d \ .$$

Expanding the binomial coefficients, we obtain (after some cancellation)

$$\frac{\binom{n}{d+i}}{\binom{n}{i}} = \frac{(n-i)(n-i-1)\ldots(n-i-d+1)}{(d+i)(d+i-1)\ldots(i+1)} = \frac{n-i}{d+i} \cdot \frac{n-i-1}{d+i-1} \cdot \ldots \cdot \frac{n-i-d+1}{i+1} \ .$$

Given our assumption that $d \leq n/5$, each of the d factors in the latter product has a value of at least 2, which accomplishes the proof.

9. If $\mathrm{SMN}'(C) \leq |X|/5$, then $\mathrm{GMN}(C) \leq 2 \cdot \mathrm{SMN}(C)$.
   Reason: This follows from inequalities that have been verified already: $\mathrm{GMN}(C) \leq \mathrm{GMN}'(C) \leq 2 \cdot \mathrm{SMN}'(C) \leq 2 \cdot \mathrm{SMN}(C)$.

3.1 Bounding AMN′ by GMN′

The verification of the inequality $\mathrm{AMN}'(C) \leq \mathrm{GMN}'(C)$ is still missing and will be given in the course of this subsection.

The (k,n)-*family* is defined as the family of concept classes C with $|C| = k$ and $|\operatorname{dom}(C)| = n$. We denote this family by $\mathcal{C}_{k,n}$.

Example 3.1. *For $k \geq 2$ and $n \geq \log k$, we denote by $C_{k,n}$ the concept class in $\mathcal{C}_{k,n}$ given by $C_{k,n} = \{c_0, c_1, \ldots, c_{k-1}\}$, with $\operatorname{dom}(C_{k,n}) = \{x_0, x_1, \ldots, x_{n-1}\}$ and $c_i(x_j)$ being the j-th bit in the binary representation of i. Formally*

$$\forall i = 0, 1, \ldots, k-1, j = 0, 1, \ldots, n-1 : c_i(x_j) = \left\lfloor \frac{i}{2^j} \right\rfloor \bmod 2$$

It is easy to characterize the $C_{k,n}$-realizable samples:

Remark 3.2. *A labeled sample S over $\operatorname{dom}(C_{k,n})$ is $C_{k,n}$-realizable iff $\sum_{j:(x_j,1)\in S} 2^j \leq k-1$.*

According to the following quite recent result, the class $C_{k,n} \in \mathcal{C}_{k,n}$ has the smallest number of realizable samples of a given size:

Theorem 3.3 (Main Theorem in [18])**.** *For each $1 \leq d \leq n$, the number of C-realizable samples of size d, with C ranging over all concept classes in $\mathcal{C}_{k,n}$, is minimized by setting $C = C_{k,n}$.*[2]

[2] The main theorem in [18] is not stated in this form, but it is equivalent to what is claimed in Theorem 3.3.

Since the minimizer $C_{k,n}$ of the number of realizable samples of size d is the same for all possible choices of d, we immediately obtain the following result:

Corollary 3.4. $\mathrm{SMN}'(C)$ *with C ranging over all concept classes in $\mathcal{C}_{k,n}$, is maximized by setting $C = C_{k,n}$.*

Here is another (less immediate than Corollary 3.4) application of Theorem 3.3:

Theorem 3.5. *For each concept class C, we have that* $\mathrm{AMN}'(C) \le \mathrm{GMN}'(C)$.

Proof. For each concept class C, let $\mathrm{AMN}''(C)$ be the smallest number d such that there exist $|C|$ many C-realizable labeled samples of size d. Since different sets of the same size always form an antichain, it follows that $\mathrm{AMN}'(C) \le \mathrm{AMN}''(C)$. It suffices therefore to show that, for each concept class C, we have that $\mathrm{AMN}''(C) \le \mathrm{GMN}'(C)$. Let C_0 be an arbitrary but fixed concept class. We set

$$k := |C_0| \ , \ \ell := \lfloor \log k \rfloor \ \text{ and } \ n := |\,\mathrm{dom}(C_0)| \ .$$

Note that $k \le 2^n$ because there can be at most 2^n distinct concepts over a domain of size n. By definition, $\mathrm{GMN}'(C_0)$ is the smallest number d such that $\sum_{i=0}^{d} \binom{n}{i} \ge k$. Note that $\mathrm{GMN}'(C)$ is the same number for all concept classes from the family $\mathcal{C}_{k,n}$. Specifically $\mathrm{GMN}'(C_0) = \mathrm{GMN}'(C_{k,n})$. On the other hand, it is immediate from Theorem 3.3 (and again the fact that different sets of the same size always form an antichain) that

$$\mathrm{AMN}''(C_{k,n}) = \max\{\mathrm{AMN}''(C) : C \in \mathcal{C}_{k,n}\} \ .$$

Thus, if we knew that $\mathrm{AMN}''(C_{k,n}) \le \mathrm{GMN}'(C_{k,n})$, we could conclude that

$$\mathrm{AMN}''(C_0) \le \mathrm{AMN}''(C_{k,n}) \le \mathrm{GMN}'(C_{k,n}) = \mathrm{GMN}'(C_0) \ .$$

It suffices therefore to verify the inequality $\mathrm{AMN}''(C_{k,n}) \le \mathrm{GMN}'(C_{k,n})$. We will make use of the following auxiliary result.

Claim 1: $\mathrm{GMN}'(C_{k,n}) \le \ell$.

Proof of Claim 1: Since $k \le 2^n$, it follows that $n \ge \lceil \log k \rceil$. The function $n \mapsto \sum_{i=0}^{\ell} \binom{n}{i}$ is monotonically increasing with n. But even for $n = \lceil \log k \rceil$, we have that

$$S := \sum_{i=0}^{\ell} \binom{\lceil \log k \rceil}{i} \ge k$$

for the following reason:

- If k is a power of 2, then $\ell = \log k = \lceil \log k \rceil$ and, therefore, $S = 2^\ell = k$.
- If k is not a power of 2, then $\ell = \lfloor \log k \rfloor$ and $\lceil \log k \rceil = \ell + 1$. This implies that $S = 2^{\ell+1} - 1 > 2^{\log k} - 1 = k - 1$ and, because S is an integer, it implies that $S \ge k$.

It follows from this discussion that $\mathrm{GMN}'(C_{k,n}) \le \ell$.

Claim 2: For each $d \in [\ell]$, there exists an injective mapping f which transforms an unlabeled sample of size at most d over domain[3] $X = \{x_0, x_1, \ldots, x_{n-1}\}$ into a (labeled) $C_{k,n}$-realizable sample of size exactly d.

Proof of Claim 2: Let $U \subseteq X$ be an unlabeled sample of size at most d. We define $f(U)$ as the (initially empty) labeled sample that is obtained from U as follows:

1. For each $x_j \in U$ with $\ell \le j \le n-1$, insert $(x_j, 0)$ into $f(U)$.
2. For each $x_j \in U$ with $0 \le j \le \ell - 1$, insert $(x_j, 1)$ into $f(U)$.
3. While $|f(U)| < d$, pick some instance x_j from $\{x_0, \ldots, x_{\ell-1}\} \setminus f(U)$ and insert $(x_j, 0)$ into $f(U)$.

It is obvious that, after Step 3, the set $f(U)$ is of size d. Moreover

$$\sum_{j:(x_j,1)\in U} 2^j \le \sum_{j=0}^{\ell-1} 2^j = 2^\ell - 1 \le k-1 \ ,$$

which, according to Remark 3.2, implies that $f(U)$ is $C_{k,n}$-realizable. Finally observe that U can be reconstructed from $f(U)$:

$$U = \{x_j : (\ell \le j \le n-1 \wedge (x_j, 0) \in f(U)) \vee (0 \le j \le \ell - 1 \wedge (x_j, 1) \in f(U)\} \ .$$

Hence f is injective, which completes the proof of Claim 2.

The proof of Theorem 3.5 can now be accomplished as follows. Set $d := \mathrm{GMN}'(C_{k,n})$. Then there exist at least k distinct unlabeled samples of size at most d over domain X. Thanks to Claim 1, we know that $d \le \ell$. Thanks to Claim 2, we may now conclude that there exist at least k distinct labeled $C_{k,n}$-realizable samples over domain X, each of size exactly d. This shows that $\mathrm{AMN}''(C_{k,n}) \le d$. □

4 Monotonicity Considerations

We say that (C', X') is an *extension of* (C, X) if $C \subseteq C'$ and $X \subseteq X'$. If $C \subsetneq C'$ and $X = X'$, it is called a *class extension*. If $X \subsetneq X'$, $C = \{c_{|X} : c \in C'\}$ and $|C| = |C'|$, it is called a *domain extension*. Note that the condition $|C| = |C'|$ ensures that X *distinguishes between the concepts in* C', i.e.,

$$\forall c^1, c^2 \in C' : c^1 \ne c^2 \Rightarrow c^1_{|X} \ne c^2_{|X} \ . \tag{2}$$

A combinatorial parameter $Y(C)$, associated with a concept class C, is said to be *class monotonic* if, for each class extension (C', X) of (C, X), we have that $Y(C) \le Y(C')$. It is said to be *domain monotonic* if, for each domain extension (C', X') of (C, X), we have that $Y(C') \le Y(C)$.

[3] We talk here about the domain of $C_{k,n}$.

	class monotonic	domain monotonic	Reference
VCD	yes	no[4]	common knowledge
RTD	yes	yes	[2]
min{VCD, RTD}	yes	no	this paper
STD_{min}	yes	yes	[12]
AMN	yes	yes	this paper
AMN′	no	yes	this paper
GMN′	yes	yes	this paper
GMN	yes	yes	this paper
SMN	yes	yes	this paper
SMN′	no	yes	this paper

Table 1: Monoconicity properties of the parameters of the SMN-GMN-AMN hierarchy.

Theorem 4.1. *The monotonicity properties of the parameters in the SMN-GMN-AMN hierarchy are as it is shown in Table 1 below.*

Proof. The monotonicity properties of VCD, RTD and STD_{min} are well known. We will now verify the remaining entries in Table 1 row by row.

min{VCD,RTD}: Set $Y = \min\{\text{VCD}, \text{RTD}\}$. For trivial reasons, the following implication is valid: if Z_1 and Z_2 are class (resp. domain) monotonic, then $\min\{Z_1, Z_2\}$ is class (resp. domain) monotonic. Setting $Z_1 = \text{VCD}$ and $Z_2 = \text{RTD}$, it follows that Y is class monotonic.

In order to show that Y is not domain monotonic, we will make use of the class C' over domain $X' = \{x_1, \ldots, x_5, x_6\}$ that is given by the following table:

C'	x_1	x_2	x_3	x_4	x_5	x_6
c_1	**1**	**1**	0	0	0	**0**
c_2	**0**	**1**	1	0	0	**0**
c_3	**0**	**0**	1	1	0	**1**
c_4	**0**	**0**	0	1	1	**0**
c_5	**1**	**0**	0	0	1	**1**
c_6	0	1	0	1	1	0
c_7	**0**	**1**	1	0	1	**1**
c_8	1	0	1	0	1	1
c_9	**1**	**0**	1	1	0	**0**
c_{10}	**1**	**1**	0	1	0	**1**

[4] VCD is domain monotonic in the other direction: for each domain extension (C', X') of (C, X), we have that $\text{VCD}(C) \leq \text{VCD}(C')$.

The class over subdomain $X = \{x_1, \ldots, x_5\}$ which is given by the first 5 columns in this table is known under the name Warmuth's class[5] and denoted by C_{MW}. Clearly (C', X') is a domain extension of (C_{MW}, X). The class C_{MW} is known to satisfy $\mathrm{VCD}(C_{MW}) = 2$ and $\mathrm{RTD}(C_{MW}) = 3$.

Claim: $\mathrm{VCD}(C') = \mathrm{RTD}(C') = 3$.

Proof: Clearly $\mathrm{VCD}(C') \leq 3$ because $\mathrm{VCD}(C) = 2$ and extending the domain by one instance (here: x_6) can increase the VC-dimension at most by 1. An inspection of the table-entries highlighted in bold reveals that $\mathrm{VCD}(C') \geq 3$. Hence $\mathrm{VCD}(C') = 3$.

Since RTD is domain monotonic, we have that $\mathrm{RTD}(C') \leq \mathrm{RTD}(C) = 3$. Assume for contradiction that $\mathrm{RTD}(C') \leq 2$. Then one of the subclasses

$$\begin{aligned} C'_0 &= \{c \in C' : c(x_6) = 0\} = \{c_1, c_2, c_4, c_6, c_9\} \\ C'_1 &= \{c \in C' : c(x_6) = 1\} = \{c_3, c_5, c_7, c_8, c_{10}\} \end{aligned}$$

must contain a concept which can be uniquely specified by one more labeled example. An inspection of the following tables for C'_0 and C'_1 reveals that this is impossible so that we arrived at a contradiction:

C'_0	x_1	x_2	x_3	x_4	x_5	x_6
c_1	1	1	0	0	0	0
c_2	0	1	1	0	0	0
c_4	0	0	0	1	1	0
c_6	0	1	0	1	1	0
c_9	1	0	1	1	0	0

C'_1	x_1	x_2	x_3	x_4	x_5	x_6
c_3	0	0	1	1	0	1
c_5	1	0	0	0	1	1
c_7	0	1	1	0	1	1
c_8	1	0	1	0	1	1
c_{10}	1	1	0	1	0	1

We conclude from this discussion that $Y(C) = 2 < 3 = Y(C')$. Thus Y is not domain monotonic.[6]

AMN: Let (C', X) be a class extension of (C, X). Let M be a C'-saturating matching of cost d that has the antichain property. Then $M_{|C}$ is a C-saturating matching of cost at most d that has the antichain property. This shows that AMN is class monotonic.
Let (C', X') be a domain extension of (C, X). Let M be a C-saturating matching of cost d that has the antichain property. Let M' be the mapping given by $M'(c') = M(c'_{|X})$. Because of (2) M' is a C'-saturating matching. Moreover M' is of the same cost as M and inherits the antichain property from M. This shows that AMN is domain monotonic.

[5]This class was designed by Manfred Warmuth so as to provide a simple class for which the RTD exceeds the VCD.

[6]It is easy to find pairs (C', X') and (C, X) such that (C', X') is a domain extension of (C, X) and $Y(C) > Y(C')$. Thus extending the domain can have effects in both directions: it can lead to a greater value of $Y = \min\{\mathrm{VCD}, \mathrm{RTD}\}$, but it can also lead to a smaller value.

AMN′: The following example shows that AMN′ is not class monotonic. Consider the concept class $C = \mathbf{C}_{16,9}$. It has 18 samples of size 1 but only 13 of them are realizable by C. Since $13 < 16 = |C|$, we get $\mathrm{AMN}'(C) \geq 2$. Let C' be the class obtained from C by adding the all-ones concept. C' is of size 17 but now all 18 samples of size 1 are realizable (and, of course, they form an antichain). Hence $\mathrm{AMN}'(C') = 1$, which is in contradiction with class monotonicity.

Let (C', X') be a domain extension of (C, X). Remember that this implies that $|C| = |C'|$. Let A be an antichain of size $|C|$ consisting of C-realizable samples of size at most d over domain X. Then A is also an antichain of size $|C'|$ consisting of C'-realizable samples of size at most d over domain X' (albeit none of these samples contains a point from $X' \setminus X$). This shows that AMN′ is domain monotonic.

GMN′: Let (C', X) be a class extension of (C, X). Then $|C'| > |C|$. Clearly $\sum_{i=0}^{d} \binom{|X|}{i} \geq |C|$ if $\sum_{i=0}^{d} \binom{|X|}{i} \geq |C'|$. Thus GMN′ is class monotonic.

Let (C', X') be a domain extension of (C, X), which implies that $|C'| = |C|$ and $|X'| > |X|$. Clearly $\sum_{i=0}^{d} \binom{|X'|}{i} \geq |C'|$ if $\sum_{i=0}^{d} \binom{|X|}{i} \geq |C|$. This shows that GMN′ is domain monotonic.

GMN: Let (C', X) be a class extension of (C, X). Let M be a greedy C-saturating matching of cost d. In order to prove class monotonicity, it suffices to specify a greedy C'-saturating matching M' of cost at least d. To this end, we consider the following C'-saturating matching M':

1. For each $c \in C$, set $M'(c) = M(c)$.
2. Inspect the concepts $c' \in C' \setminus C$ one by one. For each fixed $c' \in C' \setminus C$, choose a shortest labeled sample S' among the ones which are consistent with c' and have not already been used for another concept before. Set $M(c') = S'$.

Clearly the cost of M' is lower-bounded by the cost of M. By construction, M' cannot be directly improved on a concept $c' \in C' \setminus C$. The same remark applies to concepts $c \in C$, because any direct improvement of M' on c would be a direct improvement of M of C, which is impossible because M is greedy. It follows from this discussion that M' is a greedy C'-saturating matching of cost at least d. This shows that GMN is class monotonic.

Let (C', X') be a domain extension of (C, X), which implies that $X \subsetneq X'$. For each $c' \in C'$, let $c = c'_{|X}$ denote the corresponding concept in C. Arbitrarily fix a linear ordering $c'_1, c'_2, c'_3 \ldots$ of the concepts in C' as well as a linear extension $S_1, S_2, S_3, \ldots$ of the partial size-ordering of all C'-realizable samples over X'. Let M' be the resulting greedy C'-saturating matching and let d' denote its cost. In order to prove domain monotonicity, it suffices show that the greedy procedure applied to C has a possible output of cost at least d'. To this end, we choose the ordering $c_1, c_2, c_3 \ldots$ for the concepts in C. Furthermore, our ordering of the C-realizable samples over X (which are also C'-realizable!) is obtained from the ordering $S_1, S_2, S_3, \ldots$ by the removal of all samples containing one or more instances from $X' \setminus X$. Let M be the resulting greedy

C-saturating matching. It is easy to see that the following holds for each $i \in [m]$: if $M(c_i) = S_j$ and $M'(c'_i) = S_{j'}$, then $j' \leq j$. This clearly implies that the cost of M ís not less than the cost d' of M'. Hence GMN is domain monotonic.

SMN: Let (C', X) be a class extension of (C, X). If M' is a C'-saturating matching of cost d', then $M = M_{|C}$ is a C-saturating matching of cost $d \leq d'$. This implies that SMN is class monotonic.
Let (C', X') be a domain extension of (C, X). If M is a C-saturating matching, then M' with $M'(c') = M(c)$ for $c = c'_{|X}$ is a C'-saturating matching of the same cost. This implies that SMN is domain monotonic.

SMN′: The same example, $C = \mathbf{C}_{16,9}$, that we used above for showing that AMN′ is not class monotonic can also be used for showing that SMN′ is not class monotonic. There are 19 samples of size at most 1 (including the empty sample) but only 14 of them are realizable by C. Since $14 < 16 = |C|$, we have that $\mathrm{SMN}'(C) \geq 2$. Let again C' be the class obtained from C by adding the all-ones concept. C' is of size 17 but now all 19 samples of size at most 1 are realizable. Hence $\mathrm{SMN}'(C') = 1$, which is in contradiction with class monotonicity.
Let (C', X') be a domain extension of (C, X). Each C-realizable sample over domain X is also a (very special) C'-realizable sample over domain X'. This implies that $\mathrm{SMN}'(C') \leq \mathrm{SMN}(C)$. Thus SMN′ is domain monotonic.

This completes the proof of Theorem 4.1. □

5 Evaluation of the Parameters on the Powerset

The combinatorial parameters that we have studied so far are now evaluated for a special concept class, namely the powerset over the domain $X = \{x_1, \ldots, x_n\}$. This class will be denoted by P_n in what follows. Note that $|P_n| = 2^n$, $\mathrm{VCD}(P_n) = n$ and every sample can be realized by P_n. It is also known that $\mathrm{RTD}(P_n) = n$ [2] and that $\mathrm{STD}_{min}(P_n) = n$ [12]. Here are some more observations:

1. $\mathrm{AMN}(P_n) = \mathrm{AMN}'(P_n) = \min\left\{d : 2^d\binom{n}{d} \geq 2^n\right\}$.
Reason: As mentioned in [12], for $d \leq 1/2$ the maximum antichain contained in the set of samples of size at most d is formed by the set of labeled samples of size exactly d, a classical result in combinatorics of finite sets. This antichain has size $2^d \cdot \binom{n}{d}$. This gives the second equation. It was furthermore shown in [12] that there exists a P_n-saturating matching which assigns to each concept a labeled sample of size $\mathrm{AMN}'(C)$. This gives the first equation.

2. For all but finitely many n, we have that $0.22 \cdot n < \mathrm{AMN}(P_n) < 0.23 \cdot n$.
This was also shown in [12].

3. $\mathrm{SMN}'(P_n) = \min\left\{d : \sum_{i=0}^{d} 2^i \binom{n}{i} \geq 2^n\right\}$.
   Reason: This is immediate from $n = |X|$, $|P_n| = 2^n$, the fact that $\sum_{i=0}^{d} 2^i \binom{n}{i}$ is the number of samples of size at most d (all of which are realizable by P_n) and the definition of the parameter SMN′.

4. $\mathrm{GMN}'(P_n) = \min\left\{d : \sum_{i=0}^{d} \binom{n}{i} \geq 2^n\right\} = n$.
   Reason: The first equation is immediate from $n = |X|$, $|P_n| = 2^n$ and the definition of the parameter GMN′. The second equation is immediate from the binomial theorem.

As for the powerset, the antichain matching number and the saturating matching number are very close relatives:

Theorem 5.1. *For all but finitely many n, we have that* $\mathrm{AMN}(P_n) - 1 \leq \mathrm{SMN}'(P_n) \leq \mathrm{SMN}(P_n) \leq \mathrm{AMN}(P_n)$.

Proof. As shown before, the inequalities $\mathrm{SMN}'(C) \leq \mathrm{SMN}(C) \leq \mathrm{AMN}(C)$ are valid even for arbitrary concept classes. It suffices therefore to show that $\mathrm{SMN}'(P_n) \geq \mathrm{AMN}(P_n) - 1$, or equivalently, that $\mathrm{AMN}(P_n) \leq \mathrm{SMN}'(P_n) + 1$. Set $D = [n]$. Since

$$\mathrm{SMN}'(P_n) = \min\left\{d \in D : \sum_{i=0}^{d} 2^i \binom{n}{i} \geq 2^n\right\} \tag{3}$$

and

$$\mathrm{AMN}(P_n) = \mathrm{AMN}'(P_n) = \min\left\{d \in D : 2^d \binom{n}{d} \geq 2^n\right\} \ , \tag{4}$$

it suffices to show that

$$2^{d+1}\binom{n}{d+1} \geq \sum_{i=0}^{d} 2^i \binom{n}{i} \ . \tag{5}$$

To this end, we define

$$\Phi_d(n) = \sum_{i=0}^{d} \binom{n}{i} \quad \text{and} \quad t(n) = \begin{cases} \lceil n/3 \rceil & \text{if } n \leq 12 \\ \lfloor n/3 \rfloor + 1 & \text{if } n > 12 \end{cases} \ .$$

As proven in [10], the term $\binom{n}{d}$ is related to $\Phi_{d-1}(n)$ as follows:

$$\forall d = 1, \ldots, t(n) : \binom{n}{d} \geq \Phi_{d-1}(n) \quad \text{and} \quad \forall d = t(n)+1, \ldots, n : \binom{n}{d} < \Phi_{d-1}(n) \ .$$

Under the assumption that $d + 1 \leq t(n)$, we get $\binom{n}{d+1} \geq \Phi_d(n)$ so that

$$2^{d+1}\binom{n}{d+1} \geq 2^{d+1}\Phi_d(n) = 2^{d+1} \cdot \sum_{i=0}^{d} \binom{n}{i} > \sum_{i=0}^{d} 2^i \binom{n}{i} \ .$$

The verification of (5) will be complete if we can justify the assumption that $d + 1 \le t(n)$. This is where asymptotics comes into play. We know that, for all but finitely many n, we have that $\mathrm{SMN}'(P_n) \le \mathrm{AMN}(P_n) < 0.23 \cdot n < n/3 < t(n)$. If follows that, for all but finitely many n, the equations (3) and (4) are still valid when we set $D = [1 : t(n) - 1]$. Hence it is justified to assume that $d + 1 \le t(n)$. □

By a straightforward, but tedious, calculation, one can verify that the inequalities claimed in Theorem 5.1 are valid for all $n \ge 15$.

Theorem 5.2. *Let $H(p) = p \cdot \log(p) - (1-p) \cdot \log(1-p)$ be the binary entropy function and let p_* be the unique solution to the equation $H(p) = 1 - 2p$ subject to $p < 1/2$, and let $0 < p_0 < p_*$. Then $\mathrm{GMN}(P_n) \le \lceil (1-p_0) \cdot n \rceil$ holds for all but finitely many n.*[7]

Proof. Fix a linear ordering on the concepts in P_n and a linear extension of the partial size-ordering on the labeled samples over X such that the resulting P_n-saturating greedy matching M has cost $\mathrm{GMN}(P_n)$. We can think of M as being built in stages: in stage $0 \le j \le \mathrm{GMN}(P_n)$, each sample of size j obtains one of the concepts in C as its M-partner. Let $\mathcal{S}_r$ (resp. $\mathcal{S}_{\le r}$) denote the set of samples of size r (resp. of size at most r). We define

$$r_* := \max\left\{ r : \sum_{j=0}^{r} 2^j \cdot \binom{n}{j} \le 2^{n-r} \right\} \enspace .$$

The proof of the theorem will be based on the following two claims.

Claim 1: Every sample in $\mathcal{S}_{\le r_*}$ gets an M-partner.

Proof of Claim 1: Let's look at the moment in which an M-partner for a labeled sample $S \in \mathcal{S}_{\le r_*}$ has to be found. The number of concepts already matched at this moment is upper-bounded by $|\mathcal{S}_{\le r_*} \setminus \{S\}| \le \sum_{j=0}^{r_*} 2^j \cdot \binom{n}{j} - 1 \le 2^{n-r_*} - 1$. Hence at least one of the at least 2^{n-r_*} many neighbors of S must still be unmatched so that S will certainly obtain an M-partner. Since this reasoning applies to any concept in $\mathcal{S}_{\le r_*}$, Claim 1 follows.

Claim 2: Suppose that r is chosen such that every sample in $\mathcal{S}_{\le r}$ gets an M-partner. Suppose that q is chosen such that

$$\sum_{i=0}^{q-1} \binom{n}{i} + \sum_{j=0}^{r} \binom{n}{j} \le \sum_{j=0}^{r} 2^j \cdot \binom{n}{j} \enspace . \tag{6}$$

Then $\mathrm{GMN}(P_n) \le n - q$. In particular $\mathrm{GMN}(P_n) \le n - r - 1$.

[7] numerical calculations reveal that $p_* > 0.17$ so that we can choose $p_0 = 0.17$ and obtain $\mathrm{GMN}(P_n) \le \lceil 0.83n \rceil$.

Proof of Claim 2: We set $d = n - q$. We have to show that every concept $c \in P_n$ obtains an M-partner during stages $1, \ldots, d$. For reasons of symmetry, it suffices to prove this for the concept $c_0 = X$ (the concept assigning label 1 to every instance in X). In the remainder of the proof of Claim 2, "sample" always means "sample of size at most d". We say a labeled sample is of *type A* if it contains only 1-labeled instances. The remaining samples are said to be of *type B*. The samples of type A are precisely the ones which are consistent with c_0. We have the following situation right after stage r:

- The number of concepts with an M-partner of type B equals the number of samples of type B within $\mathcal{S}_{\le r}$, and the latter number equals $\sum_{j=0}^{r} 2^j \binom{n}{j} - \sum_{j=0}^{r} \binom{n}{j}$.
- $E_1 := 2^n - \left(\sum_{j=0}^{r} 2^j \binom{n}{j} - \sum_{j=0}^{r} \binom{n}{j} \right)$ is the number of the remaining concepts.

The number of samples of type A clearly equals

$$E_2 := \sum_{j=0}^{d} \binom{n}{j} = 2^n - \sum_{j=d+1}^{n} \binom{n}{j} = 2^n - \sum_{j=0}^{n-d-1} \binom{n}{j} = 2^n - \sum_{j=0}^{q-1} \binom{n}{j} \ . \tag{7}$$

Note that $E_1 \le E_2$ is equivalent to (6). Since q, by assumption, is chosen such that the condition (6) is satisfied, we may conclude that $E_1 \le E_2$. What does $E_1 \le E_2$ mean? It means that the number of concepts in $P_n \setminus \{c_0\}$ not being matched with a labeled sample of type B is smaller than the number of samples of type A. Since c_0 is adjacent to all samples of type A, this implies that c_0 will obtain an M-partner in stage d (or even in an earlier stage). Hence $\mathrm{GMN}(P_n) \le d = n - q$, Finally note that (6) can be satisfied by setting $q = r + 1$. Hence $\mathrm{GMN}(P_n) \le n - q - 1$, which completes the proof of Claim 2.

Let r be a number with the property that

$$r + 1 \le \frac{n}{3} \text{ and } 2^{r+2} \cdot \binom{n}{r+1} \le 2^{n-r} \ .$$

The condition $r + 1 \le n/3$ implies that $\binom{n}{r+1} \ge \sum_{j=0}^{r} \binom{n}{j}$ so that $2^{r+2} \cdot \binom{n}{r+1} > \sum_{j=0}^{r} 2^j \cdot \binom{n}{j}$. An inspection of the definition of r_* reveals that $r \le r_*$. We can therefore infer from Claim 1 that every sample in $\mathcal{S}_{\le r}$ gets an M-partner. Setting $p := \frac{r+1}{n}$. we obtain the following series of equivalent conditions:

$$2^{r+2} \cdot \binom{n}{r+1} \le 2^{n-r} \Leftrightarrow \binom{n}{r+1} \le 2^{n-2r-2} \Leftrightarrow \binom{n}{pn} \le 2^{n-2pn} \Leftrightarrow \frac{1}{n} \log \binom{n}{pn} \le 1 - 2p \ .$$

It is well known that $\frac{1}{n} \log \binom{n}{pn}$ converges to $H(p)$ if n goes to infinity. Let $p_* < 1/2$ be the unique solution of the equation $H(p) = 1 - 2p$. Fix any constant $0 < p_0 < p_*$. The theorem can now be obtained as follows. The inequality $p_0 < p_*$ implies that $\frac{1}{n} \log \binom{n}{p_0 n} < 1 - 2 \cdot p_0$ provided that n is sufficiently large. This in turn implies that, for r_0' given by $r_0' + 1 = p_0 \cdot n$, we have that $2^{r'_0+2} \cdot \binom{n}{r'_0+1} < 2^{n-r'_0}$. Setting $r_0 = \lfloor r'_0 \rfloor$, we get $2^{r_0+2} \cdot \binom{n}{r_0+1} < 2^{n-r_0}$. According to Claim 2, we have

$$\mathrm{GMN}(P_n) \le n - r_0 - 1 = n - \lfloor r'_0 \rfloor - 1 = n - \lfloor r'_0 + 1 \rfloor = n - \lfloor p_0 n \rfloor = \lceil (1-p_0)n \rceil \ .$$

This completes the proof of Theorem 5.2. □

6 Additivity Considerations

Definition 6.1. *Suppose that C_1 is a concept class over domain X_1, C_2 is a concept class over domain X_2 and $X_1 \cap X_2 = \emptyset$. The* free combination *of C and C', denoted by $C_1 \sqcup C_2$, is then given by*

$$C_1 \sqcup C_2 := \{c_1 \cup c_2 : c_1 \in C_1 \text{ and } c_2 \in C_2\} \ .$$

A combinatorial parameter Y is said to be sub-additive *(resp.* super-additive*) if*

$$Y(C_1 \sqcup C_2) \le Y(C_1) + Y(C_2) \qquad (8)$$

(resp. $Y(C_1 \sqcup C_2) \ge Y(C_1) + Y(C_2)$) holds for each free combination of two concept classes C_1 and C_2. If (8) holds with equality for each free combination of two concept classes, then Y is said to be additive*,*

The goal of this section is to verify the entries of Table 2 below.

	additive	sub-additive	super-additive	Reference
VCD	yes	yes	yes	common knowledge
RTD	yes	yes	yes	[2]
$\min\{\mathrm{VCD}, \mathrm{RTD}\}$	no	no	yes	this paper
STD_{min}	no	yes	no	this paper
AMN	no	yes	no	this paper
AMN′	no	yes	no	this paper
GMN′	no	yes	no	this paper
GMN	no	?	no	this paper
SMN	no	yes	no	this paper
SMN′	no	yes	no	this paper

Table 2: Additivity properties of the parameters of the SMN-GMN-AMN hierarchy.

Remark 6.2. VCD *and* RTD *are known to be additive.*

The following relation, which holds for arbitrary reals a, b, c, d is well known and easy to check:

$$\min\{a+b, c+d\} \ge min\{a,c\} + \min\{b,d\} \ . \qquad (9)$$

This inequality is strict if $a < c$ and $b > d$. Moreover, we say that two parameters Y, Z are *incomparable* if there exist C, C' with $Y(C) < Z(C)$ and $Y(C') > Z(C')$

Remark 6.3. *1. If Z_1 and Z_2 are super-additive, then $Y := \min\{Z_1, Z_2\}$ is super-additive too.*

2. If Z_1 and Z_2 are additive and incomparable, then $Y := \min\{Z_1, Z_2\}$ is not additive.

Proof. Choose any two concept classes C_1 and C_2 over disjoint domains. Then:

$$\begin{aligned} Y(C_1 \sqcup C_2) &= \min\{Z_1(C_1 \sqcup C_2), Z_2(C_1 \sqcup C_2)\} \\ &\geq \min\{Z_1(C_1) + Z_1(C_2), Z_2(C_1) + Z_2(C_2)\} \\ &\overset{(*)}{\geq} \min\{Z_1(C_1), Z_2(C_1)\} + \min\{Z_1(C_2), Z_2(C_2)\} = Y(C_1) + Y(C_2)\ . \end{aligned}$$

The first equation holds by definition of Y. The subsequent inequality holds by the super-additivity of Z_1 and Z_2. The inequality which is marked "$(*)$" is an application of (9). The final equation holds by definition of Y again. The above calculation shows that Y is super-additive. Note that the inequality which is marked "$(*)$" becomes strict if C_1 and C_2 are chosen such that $Z_1(C_1) < Z_2(C_1)$ and $Z_2(C_2) < Z_1(C_2)$. Note that this is a possible choice under the assumption that Z_1 and Z_2 are incomparable. Thus, if Z_1 and Z_2 are incomparable, then Y is not additive. □

Example 6.4. *We may now conclude from the additivity and incomparability of* VCD *and* RTD *(this is well-known) that* $\min\{\mathrm{VCD}, \mathrm{RTD}\}$ *is super-additive but not additive.*

Remark 6.5. STD_{min} *is sub-additive but not additive.*

Proof. Choose any two concept classes C_1 and C_2 over disjoint domains $X_1 = \mathrm{dom}(C_1)$ and $X_2 = \mathrm{dom}(C_2)$. For $i = 1, 2$, let $(M_k^i)_{k=0,\ldots,k^*(i)}$ be a subset teaching sequence of cost $d_i := \mathrm{STD}_{min}(C_i)$ for C_i. Let M_{k_1,k_2} be the $(C_1 \sqcup C_2)$-saturating matching given by $M_{k_1,k_2}(c_1 \cup c_2) = M_{k_1}(c_1) \cup M_{k_2}(c_2)$. It is easy to see that the sequence which starts with $M_{0,0}, M_{1,0}1, \ldots, M_{k^*(1),0}$ and continues with $M_{k^*(1),0}, M_{k^*(1),1}, \ldots, M_{k^*(1),k^*(2)}$ is a subset teaching sequence of cost $d_1 + d_2$ for $C_1 \sqcup C_2$. Hence STD_{min} is sub-additive. Let now $C_1 = \{c_0, c_1, c_2, c_3\}$ be the concept class over domain $X_1 = \{u, v\}$ given by $c_0 = \emptyset$, $c_1 = \{u\}$, $c_2 = \{v\}$ and $c_3 = \{u, v\}$. Let C_2 consist of the single concept X_2 over domain $X_2 = \{a, b, c, d\}$. Then $\mathrm{STD}_{min}(C_1) = 2$ and $\mathrm{STD}_{min}(C_2) = 0$. The following subset teaching sequence shows that $\mathrm{STD}_{min}(C \sqcup C') = 1$:

$$\begin{array}{rcl} M_0(c_0 \cup X_2) &=& (u,0),(v,0),(a,1),(b,1),(c,1),(d,1) \\ M_0(c_1 \cup X_2) &=& (u,1),(v,0),(a,1),(b,1),(c,1),(d,1) \\ M_0(c_2 \cup X_2) &=& (u,0),(v,1),(a,1),(b,1),(c,1),(d,1) \\ M_0(c_1 \cup X_2) &=& (u,1),(v,1),(a,1),(b,1),(c,1),(d,1) \\ \hline M_2(c_0 \cup X_2) &=& (u,0),(v,0),(a,1) \\ M_2(c_1 \cup X_2) &=& (u,1),(v,0),(b,1) \\ M_2(c_2 \cup X_2) &=& (u,0),(v,1),(c,1) \\ M_2(c_1 \cup X_2) &=& (u,1),(v,1),(d,1) \\ \hline M_3(c_0 \cup X_2) &=& (a,1) \\ M_3(c_1 \cup X_2) &=& (b,1) \\ M_3(c_2 \cup X_2) &=& (c,1) \\ M_3(c_1 \cup X_2) &=& (d,1) \end{array}$$

It follows that STD_{min} is not additive. □

The powerset P_n over the domain $X = \{x_1, \ldots, x_n\}$ can be seen as the free combination of $P_1^1 \sqcup \ldots \sqcup P_1^n$ where P_1^i is the powerset over the domain $\{x_i\}$. For trivial reasons, we have the following implications:

- If Y is additive, then $Y(P_n) = n \cdot Y(P_1)$.
- If Y is sub-additive, then $Y(P_n) \leq n \cdot Y(P_1)$.
- If Y is super-additive, then $Y(P_n) \geq n \cdot Y(P_1)$.

Example 6.6. *It is easy to check that all parameters in the SMN-GMN-AMN hierarchy evaluate to* 1 *on the class* P_1*. We know already that* $\mathrm{GMN}(P_n)$ *and* $\mathrm{AMN}(P_n)$ *are less than* n *provided that* n *is sufficiently large. Of course, the same must be true for all parameters below* GMN *and* AMN *in the SMN-GMN-AMN hierarchy. We may conclude from this short discussion that none of the parameters* $\mathrm{GMN}, \mathrm{SMN}, \mathrm{SMN}', \mathrm{AMN}, \mathrm{AMN}'$ *is super-additive (which implies that none of them is additive).*

Theorem 6.7. *Each of the parameters* $\mathrm{GMN}', \mathrm{SMN}, \mathrm{SMN}', \mathrm{AMN}, \mathrm{AMN}'$ *is sub-additive but not additive.*

Proof. Choose any two concept classes C_1 and C_2 over disjoint domains X_1 and X_2, respectively. Clearly

$$|C_1 \sqcup C_2| = |C_1| \cdot |C_2| \ .$$

This multiplicativity makes it easy to show sub-additivity for various parameters:

1. We first consider the parameter GMN'. For $i = 1, 2$, let $U_i = \binom{X_i}{\leq d_i}$ be the set of unlabeled samples of size at most d_i with sample points taken from X_i. Let $U = \binom{X_1 \cup X_2}{\leq d_1 + d_2}$ be the set of unlabeled samples of size at most $d_1 + d_2$ with sample points taken from $X_1 \cup X_2$. Then $(S_1, S_2) \mapsto S_1 \cup S_2$ is an injective mapping from $U_1 \times U_2$ to U. Thus $|U| \geq |U_1| \cdot |U_2|$. Hence, if $|U_1| \geq |C_1|$ and $|U_2| \geq |C_2|$, then

   $$|U| \geq |U_1| \cdot |U_2| \geq |C_1| \cdot |C_2| = |C_1 \sqcup C_2| \ ,$$

   which implies the sub-additivity of GMN'.

2. Let us now consider the parameter SMN'. The reasoning is analogous to the reasoning for the parameter GMN'. This time U_i is the number of C_i-realizable (labeled) samples of size at most d_i and U is the set of $(C_1 \sqcup C_2)$ realizable samples of size at most $d_1 + d_2$. The main observations is that the union of a C_1-realizable sample with a C_2 realizable sample yields a $(C_1 \sqcup C_2)$-realizable sample so that $(S_1, S_2) \mapsto S_1 \cup S_2$ is an injective mapping from $U_1 \times U_2$ to U. Moreover, if M_i is a C_i-saturating matching in (C_i, X_i) of cost d_i (for $i = 1, 2$), then M given by $M(c_1 \cup c_2) = M_1(c_1) \cup M_2(c_2)$ is a $(C_1 \sqcup C_2)$-saturating matching. It follows that SMN' and SMN are sub-additive.

3. The sub-additivity of the parameters AMN' and AMN can be shown in a similar fashion. The only observation that we need, in addition to the previous ones, is the following: if $U_i \subseteq 2^{X_i \times \{0,1\}}$ is a collection of samples which forms an antichain (for $i = 1, 2$), then $U = \{S_1 \cup S_2 : S_1 \in U_1 \text{ and } S_2 \in U_2\}$ is an antichain too.

We mentioned already in Example 6.6 that the parameters $\mathrm{SMN}, \mathrm{SMN}', \mathrm{AMN}, \mathrm{AMN}'$ are not additive. In order to complete the verification of the entries in Table 2, we still have to show that GMN' is not additive. Choose both concept classs, C_1 and C_2, from the (k, n)-family with $k = n+2$, i.e., $|X_i| = n$ and $|C_i| = n+2$ for $i = 1, 2$. Then $\mathrm{GMN}'(C_1) = \mathrm{GMN}'(C_2) = 2$. A simple calculation shows that

$$\binom{2n}{0} + \binom{2n}{1} + \binom{2n}{2} + \binom{2n}{3} > (n+2)^2 = |C_1 \sqcup C_2|$$

holds for all $n \geq 3$. Hence, assuming $n \geq 3$, we have

$$\mathrm{GMN}'(C_1 \sqcup C_2) \leq 3 < 4 = \mathrm{GMN}'(C_1) + \mathrm{GMN}'(C_2) \ ,$$

which shows that GMN' is not additive. This complertes the proof of Theorem 6.7. □

7 A Final Look at the SMN-GMN-AMN Hierarchy

A *combinatorial parameter (associated with concept classes)* is a mapping $C \to Y(C)$ which assigns a non-negative number $Y(C)$ to each (finite) concept class C. Suppose that Y and Z are two combinatorial parameters. We will write $Y \to Z$ if the following hold:

1. For each concept class C, we have that $Y(C) \leq Z(C)$.

2. The $\leq$-relation between $Y(C)$ and $Z(C)$ is *occasionally strict*, i.e., there exists a concept class C such that $Y(C) < Z(C)$.

We say that Y and Z are *incomparable* if there exist concept classes C_1 and C_2 such that $Y(C_1) < Z(C_1)$ and $Z(C_2) < Y(C_2)$. This is the case iff $Y \neq Z$ but neither $Y \to Z$ nor $Z \to Y$.

In the course of this paper, we have verified various $\leq$-relations between combinatorial parameters associated with concept classes. We will show now, by a series of examples and remarks, that all these $\leq$-relations are occasionally strict so that we arrive at the diagram in Fig. 1 below.

Remark 7.1. *Suppose that $Y(C) \leq Z(C)$ for each concept class C. Then, if Z is class monotonic but Y is not, then $Y \to Z$. Specifically, if (C', X) is a class extension of (C, X) such that $Y(C') < Y(C)$, then $Y(C') < Z(C')$.*

Proof. Let (C', X) be a class extension of (C, X) such that $Y(C') < Y(C)$. Then $Z(C') \geq Z(C) \geq Y(C) > Y(C')$. □

Example 7.2. *We know already that* SMN *is class monotonic while* SMN′ *is not. Specifically, let* $C = \mathbf{C}_{16,9}$ *and let* C' *be obtained from* C *by adding the all-ones concept. Then, as we have already shown before, we have that* $\mathrm{SMN}(C') = 1 < \mathrm{SMN}(C')$. *Now an application of Remark 7.1 yields* $\mathrm{SMN}'(C') < \mathrm{SMN}(C')$. *Hence* $\mathrm{SMN}' \to \mathrm{SMN}$.

Example 7.3. *Clearly* $\mathrm{SMN}(P_2) \geq 1$. *The matching* M *with*

$$M(\emptyset) = \emptyset\ , M(\{x_1\}) = \{(x_2, 0)\}\ ,\ M(\{x_2\}) = \{(x_2, 1)\}\ ,\ M(\{x_1, x_2\}) = \{(x_1, 1)\}$$

witnesses that $\mathrm{SMN}(P_2) \leq 1$. *Thus* $\mathrm{SMN}(P_2) = 1$. *Clearly* $\mathrm{GMN}(P_2) \leq 2$. *Consider the ordering* $\{x_1, x_2\}, \{x_1\}, \{x_2\}, \emptyset$ *for the concepts in* P_2 *and the ordering*

$$\emptyset, \{(x_2, 0)\}, \{(x_1, 0)\}, \{(x_1, 1)\}, \{(x_2, 1)\}, \{(x_1, 0), (x_2, 0)\} \ldots$$

for the labeled samples over the domain $\{x_1, x_2\}$. *The resulting greedy matching* M' *is then given by*

$$M'(\{x_1, x_2\} = \emptyset\ ,\ M'(\{x_1\}) = \{(x_2, 0)\}\ ,\ M'(\{x_2\} = \{(x_1, 0)\}\ ,\ M'(\emptyset) = \{(x_1, 0), (x_2, 0)\}\ .$$

It follows that $\mathrm{GMN}(P_2) \geq 2$. *Thus* $\mathrm{GMN}(P_2) = 2 > 1 = \mathrm{SMN}(P_2)$ *and, therefore,* $\mathrm{SMN} \to \mathrm{GMN}$.

Remark 7.4. *We know from Theorem 5.2 that* $\mathrm{GMN}(P_n) < n$ *for all sufficiently large* n. *We know furthermore that* $\mathrm{GMN}'(P_n) = n$ *for all* $n \geq 1$. *Hence* $\mathrm{GMN} \to \mathrm{GMN}'$.

Remark 7.5. *Suppose that* Y, Z, Z' *are combinatorial parameters where* Z *and* Z' *are incomparable and, for each concept class* C, *we have that* $Y(C) \leq Z(C)$ *and* $Y(C) \leq Z'(C)$. *Then* $Y \to Z$ *and* $Y \to Z'$.

Proof. Pick concept classes C and C' such that $Z(C) < Z'(C)$ and $Z'(C') < Z(C')$. Then $Y(C) \leq Z(C) < Z(C')$ and $Y(C') \leq Z'(C') < Z(C')$. Hence $Y \to Z'$ and $Y \to Z$. □

Example 7.6. *Consider concept class* $C = \mathbf{C}_{k,n}$ *with* $k = 2^\ell$ *for some* $\ell \geq 2$ *and* $n \geq k - 1$. *Since* $\binom{n}{0} + \binom{n}{1} = 1 + n \geq k$, *it follows that* $\mathrm{GMN}'(C) = 1$. *It is easy to see that* $\mathrm{VCD}(C) = \mathrm{RTD}(C) = \ell$. *It follows that* $\mathrm{GMN}'(C) < \min\{\mathrm{VCD}(C), \mathrm{RTD}(C)\}$ *and, therefore,* $\mathrm{GMN}' \to \min\{\mathrm{VCD}, \mathrm{RTD}\}$.

Example 7.7. *The smallest number* d *such that* $\sum_{i=0}^{d} 2^i \binom{6}{i} \geq 2^6$ *equals* 2 *whereas the smallest number* d *such that* $2^d \binom{6}{d} \geq 2^6$ *equals* 3. *It follows that* $\mathrm{SMN}'(P_6) = 2 < 3 = \mathrm{AMN}'(P_6)$. *Thus* $\mathrm{SMN}' \to \mathrm{AMN}'$.

Example 7.8. *We know already that* AMN *is class monotonic while* AMN′ *is not. Specifically, let* $C = \mathbf{C}_{16,9}$ *and let* C' *be obtained from* C *by adding the all-ones concept. Then, as we have already shown before, we have that* $\mathrm{AMN}'(C') = 1 < \mathrm{AMN}'(C)$. *Now an application of Remark 7.1 yields* $\mathrm{AMN}'(C') < \mathrm{AMN}(C')$. *Hence* $\mathrm{AMN}' \to \mathrm{AMN}$.

Remark 7.9. *We mentioned already in Section 5 that* $\mathrm{STD}_{min}(P_n) = n$ *is valid for all* $n \geq 1$ *and* $\mathrm{AMN}(P_n) < 0.23n$ *is valid for all sufficiently large* n. *It follows that* $\mathrm{AMN} \to \mathrm{STD}_{min}$.

Remark 7.10. *Suppose that Z_1 and Z_2 are incomparable additive combinatorial parameters and Y is a sub-additive combinatorial parameter which satisfies $Y(C) \le Z_1(C)$ and $Y(C) \le Z_2(C)$ for each concept class C. Then $Y \to \min\{Z_1, Z_2\}$.*

Proof. Pick two concept classes C_1 and C_2 such that $Z_1(C_1) < Z_2(C_1)$ and $Z_2(C_2) < Z_1(C_2)$. Rename the domain elements (if necessary) such that $\mathrm{dom}(C_1) \cap \mathrm{dom}(C_2) = \emptyset$. Then

$$\begin{aligned} Y(C_1 \sqcup C_2) &\le Y(C_1) + Y(C_2) \le Z_1(C_1) + Z_2(C_2) \\ &< \min\{Z_1(C_1) + Z_1(C_2), Z_2(C_1) + Z_2(C_2)\} = \min\{Z_1(C_1 \sqcup C_2), Z_2(C_1 \sqcup C_2)\} \ . \end{aligned}$$

It follows that $Y \to \min\{Z_1, Z_2\}$. □

Example 7.11. *It is well known that* VCD *and* RTD *are additive. According to Remark 6.5,* STD_{min} *is sub-additive. By an application of Remark 7.10, we get* $\mathrm{STD}_{min} \to \min\{\mathrm{VCD}, \mathrm{RTD}\}$.

Example 7.12. *We know already that* $\mathrm{AMN}(P_6) = \mathrm{AMN}'(P_6) = 3$. *We claim* $\mathrm{SMN}(P_6) = 2$, *which would imply that* $\mathrm{SMN} \to \mathrm{AMN}$. $\mathrm{SMN}(P_6) = 2$ *can be shown by matching the empty set with the empty sample, a singleton set* $\{i\}$ *by the sample* $\{(i, 1)\}$, *a co-singleton set* $[6] \setminus \{i\}$ *by the sample* $\{(i, 0)\}$ *and by matching the remaining concepts in* P_6 *by appropriately chosen samples of size* 2. *We need to argue that such a saturating matching for the remaining concepts does exist. To this end consider the bipartite consistency graph (the graph with one vertex for each concept* c, *one vertex for each labeled sample* S *and an edge between them iff* c *is consistent with* S. *Each concept is consistent with* $\binom{6}{2} = 15$ *samples of size* 2. *Each sample of size* 2 *is consistent with* $2^4 = 16$ *concepts from* P_6, *including the empty set as well as the singleton and co-singleton sets. But, if we exclude these sets, then each sample of size* 2 *is consistent with at most* 10 *of the remaining concepts. Now it is an easy application of Hall's theorem to show that all remaining concepts can be matched with appropriately chosen samples of size* 2.

Remark 7.13. *We mentioned already in Section 5 that* $\mathrm{GMN}'(P_n) = n$ *is valid for all* $n \ge 1$ *and* $\mathrm{AMN}(P_n) < 0.23n$ *is valid for all sufficiently large* n. *Since* $\mathrm{AMN}'(P_n) = \mathrm{AMN}(P_n)$, *it follows that* $\mathrm{AMN}' \to \mathrm{GMN}'$.

This concludes the verification of the diagram in Fig. 1.

Open Problem. We have left open the question of whether GMN is sub-additive.

References

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

[2] 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.

[3] Shaun Fallat, David Kirkpatrick, Hans U Simon, Abolghasem Soltani, and Sandra Zilles. On batch teaching without collusion. *Journal of Machine Learning Research*, 24:1–33, 2023.

[4] Cèsar Ferri, Darío Garigliotti, José Hernández-Orallo, Brigt Arve Toppe Håvardstun, and Jan Arne Telle. When redundancy matters: Machine teaching of representations. *Mach. Learn.*, 115(1):20, 2026.

[5] Cèsar Ferri, José Hernández-Orallo, and Jan Arne Telle. Non-cheating teaching revisited: A new probabilistic machine teaching model. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 2973–2979. ijcai.org, 2022.

[6] Sally Floyd and Manfred K. Warmuth. Sample compression, learnability, and the vapnik-chervonenkis dimension. *Mach. Learn.*, 21(3):269–304, 1995.

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

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

[9] Lunjia Hu, Ruihan Wu, Tianhong Li, and Liwei Wang. Quadratic upper bound for recursive teaching dimension of finite VC classes. In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, Proceedings of Machine Learning Research, pages 1147–1156. PMLR, 2017.

[10] Norbert Klasner and Hans U. Simon. General lower bounds on the query complexity within the exact learning model. *Discrete Applied Mathematics*, 107(1–3):61–81, 2000.

[11] Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Mach. Learn.*, 2(4):285–318, 1987.

[12] Farnam Mansouri, Hans U. Simon, Adish Singla, and Sandra Zilles. On batch teaching with sample complexity bounded by vcd. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, 2022.

[13] Rahim Samei, Pavel Semukhin, Boting Yang, and Sandra Zilles. Algebraic methods proving sauer's bound for teaching. *Theoretical Computer Science*, 558:35–50, 2014.

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

[15] 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.

[16] S. Shelah. A combinatorial problem: Stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41:247–261, 1972.

[17] Hans Ulrich Simon and Jan Arne Telle. Map- and mle-based teaching. *Journal of Machine Learning Research*, 25(96), 2024.

[18] Joakim Sunde, Brigt Håvardstun, Jan Kratochvíl, and Jan Arne Telle. On a combinatorial problem arising in machine teaching. In *Proceedings of the 41st International Conference on Machine Learning*, pages 47305–47313, 2024.

[19] Leslie G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

[20] 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.

[21] Chaoqi Wang, Adish Singla, and Yuxin Chen. Teaching an active learner with contrastive examples. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 17968–17980, 2021.

[22] Scott Cheng-Hsin Yang, Wai Keen Vong, Ravi B Sojitra, Tomas Folke, and Patrick Shafto. Mitigating belief projection in explainable artificial intelligence via bayesian teaching. *Scientific reports*, 11(1):1–17, 2021.

[23] Xuezhou Zhang, Shubham Kumar Bharti, Yuzhe Ma, Adish Singla, and Xiaojin Zhu. The sample complexity of teaching by reinforcement on q-learning. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 10939–10947. AAAI Press, 2021.

[24] Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N Rafferty. An overview of machine teaching. *arXiv preprint arXiv:1801.05927*, 2018.

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