Stability Is Stable: Connections between Replicability, Privacy, and Adaptive Generalization*

Mark Bun
mbun@bu.edu
Boston University
Boston, MA, USA

Marco Gaboardi
gaboardi@bu.edu
Boston University
Boston, MA, USA

Max Hopkins
nmhopkin@eng.ucsd.edu
University of California San Diego
San Diego, CA, USA

Russell Impagliazzo
russell@eng.ucsd.edu
University of California San Diego
San Diego, CA, USA

Rex Lei
rlei@ucsd.edu
University of California San Diego
San Diego, CA, USA

Toniann Pitassi
tonipitassi@gmail.com
Columbia University
New York, New York, USA

Satchit Sivakumar
satchit@bu.edu
Boston University
Boston, MA, USA

Jessica Sorrell
jsorrell@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

ABSTRACT

The notion of replicable algorithms was introduced by Impagliazzo, Lei, Pitassi, and Sorrell (STOC'22) to describe randomized algorithms that are stable under the resampling of their inputs. More precisely, a replicable algorithm gives the same output with high probability when its randomness is fixed and it is run on a new i.i.d. sample drawn from the same distribution. Using replicable algorithms for data analysis can facilitate the verification of published results by ensuring that the results of an analysis will be the same with high probability, even when that analysis is performed on a new data set.

In this work, we establish new connections and separations between replicability and standard notions of algorithmic stability. In particular, we give sample-efficient algorithmic reductions between perfect generalization, approximate differential privacy, and replicability for a broad class of statistical problems. Conversely, we show any such equivalence must break down computationally: there exist statistical problems that are easy under differential privacy, but that cannot be solved replicably without breaking public-key cryptography. Furthermore, these results are tight: our reductions are statistically optimal, and we show that any computational separation between DP and replicability must imply the existence of one-way functions.

Our statistical reductions give a new algorithmic framework for translating between notions of stability, which we instantiate to answer several open questions in replicability and privacy. This includes giving sample-efficient replicable algorithms for various PAC learning, distribution estimation, and distribution testing problems, algorithmic amplification of δ in approximate DP, conversions from item-level to user-level privacy, and the existence of private agnostic-to-realizable learning reductions under structured distributions.

CCS CONCEPTS

• **Theory of computation** → **Machine learning theory**; **Algorithm design techniques**.

KEYWORDS

Replicability, Differential Privacy, Generalization, Algorithmic Stability

ACM Reference Format:
Mark Bun, Marco Gaboardi, Max Hopkins, Russell Impagliazzo, Rex Lei, Toniann Pitassi, Satchit Sivakumar, and Jessica Sorrell. 2023. Stability Is Stable: Connections between Replicability, Privacy, and Adaptive Generalization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23), June 20–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3564246.3585246

1 INTRODUCTION

Replicability is the principle that the findings of an empirical study should remain the same when it is repeated on new data. Despite being a pillar of the scientific method, replicability is extremely difficult to ensure in today's complex data generation and analysis processes. Questionable research practices including misapplication of statistics, selective reporting of only the findings that appear most statistically significant, and the formulation of research hypotheses after the results are already known have been identified as causes of an ongoing "crisis of replicability" across the empirical sciences. Toward formulating solutions in the context of machine learning and algorithmic data analysis, Impagliazzo, Lei, Pitassi,

*All section, theorem, and lemma references are to the full version on arXiv (v2). The detailed statements and proofs of all results can be found there.

STOC '23, June 20–23, 2023, Orlando, FL, USA

ACM ISBN 978-1-4503-9913-5/23/06…$15.00
https://doi.org/10.1145/3564246.3585246

and Sorrell [23] recently put forth a new definition of replicability for statistical learning algorithms.[1]

Definition 1.1. *A randomized algorithm $A : \mathcal{X}^n \to \mathcal{Y}$ is ρ-replicable if for every distribution D over $\mathcal{X}$, we have*

$$\mathbf{Pr}[A(S_1; r) = A(S_2; r)] \geq 1 - \rho,$$

where $S_1, S_2 \in \mathcal{X}^n$ are independent sequences of i.i.d. samples from D, and r represents the coin tosses of the algorithm A.

That is, an algorithm (capturing an end-to-end data analysis process) is replicable if with high probability over the choice of two independent samples from the same distribution, it produces exactly the same output. If one research team shares both their replicable analysis process (A) and the random choices made along the way (r), then another research team can independently verify their conclusions by performing the same analysis on a fresh dataset.

Replicability is an extremely strong *stability* constraint to place on an algorithm. Informally, an algorithm is stable if its output is insensitive to small changes to its input. Nevertheless, replicability is achievable for many fundamental data analysis tasks, including statistical query learning, heavy hitter identification, approximate median finding, and large-margin halfspace learning [20, 23].

Replicability is not the first definition of algorithmic stability aimed at ensuring the utility and safety of modern data analysis. Others have played central roles in relatively mature areas such as differential privacy and adaptive data analysis. Some of the aforementioned replicable algorithms were, in fact, motivated or inspired by differentially private counterparts. Is there a systematic explanation for this? *What can we learn about the capabilities and limitations of replicable algorithms by relating replicability to other notions of algorithmic stability?*

Let us briefly recall the types of algorithmic stability that arise in these other areas:

Differential privacy. A randomized algorithm is differentially private [16] if changing a single input record results in a small change in the distribution of the algorithm's output. When each input record corresponds to one individual's datum, differential privacy guarantees that nothing specific to any individual can be learned from the output of the algorithm. (See Section 2.4.) Differential privacy comes with a rich algorithmic toolkit and understanding of the feasibility of fundamental statistical tasks in query estimation, classification, regression, distribution estimation, hypothesis testing, and more.

Generalization in adaptive data analysis. Generalization is the ability of a learning algorithm to reflect properties of a population, rather than just properties of a specific sample drawn from that population. Techniques for provably ensuring generalization form a hallmark of theoretical machine learning. However, generalization is particularly difficult to guarantee in settings where multiple analyses are performed adaptively on the same sample. Traditional notions of generalization do not hold up to downstream misinterpretation of results. For example, a classifier that encodes detailed information about its training sample in its lower order bits may generalize well, but can be used to construct a different classifier that behaves very differently on the sample than it does on the population. Interactive processes such as exploratory data analysis or feature selection followed by classification/regression can ruin the independence between the training sample and the method used to analyze it, invalidating standard generalization arguments.

Adaptivity in data analysis has been identified as one contributing factor to the replication crisis, and imposing stability conditions on learning algorithms offers solutions to this part of the problem. A variety of such stability conditions have been studied [4–6, 12, 14, 15, 27, 30, 34, 36], each offering distinct advantages in terms of the breadth of their applicability and the quantitative parameters achievable. Two specific notions play a central role in this work. The first is *perfect generalization* [4, 12], which ensures that whatever can be inferred from the output of a learning algorithm when run on a sample S could have been learned just from the underlying population itself:

Definition 1.2. *An algorithm $A : \mathcal{X}^n \to \mathcal{Y}$ is $(\beta, \varepsilon, \delta)$-perfectly generalizing if, for every distribution D over $\mathcal{X}$, there exists a distribution Sim_D such that, with probability at least $1 - \beta$ over S consisting of n i.i.d. samples from D, and every set of outcomes $O \subseteq \mathcal{Y}$,*

$$e^{-\varepsilon}(\mathbf{Pr}_{Sim_D}[O] - \delta) \leq \mathbf{Pr}[A(S) \in O] \leq e^{\varepsilon}\mathbf{Pr}_{Sim_D}[O] + \delta. \quad (1)$$

The second is *max-information* [14] which constrains the amount of information revealed to an analyst about the training sample:

Definition 1.3. *An algorithm $A : \mathcal{X}^n \to \mathcal{Y}$ has (ε, δ)-max-information with respect to product distributions if for every set of outcomes $O \subseteq (\mathcal{Y} \times \mathcal{X}^n)$ we have*

$$\mathbf{Pr}[(A(S), S) \in O] \leq e^{\varepsilon}\mathbf{Pr}[(A(S), S') \in O] + \delta,$$

where S and S' are independent samples of size n drawn i.i.d. from an arbitrary distribution D over $\mathcal{X}$.

As with differential privacy, both perfect generalization and max-information are robust to post-processing.

Each stability definition described above is tailored to model a distinct desideratum. At first glance, they may all appear technically incomparable. For instance, differential privacy is stricter than the other definitions in that it holds in the worst case over all input datasets without any assumptions on the data-generating procedure. On the other hand, it is weaker in that it only requires insensitivity to changing one input record, rather than to resampling the entire input dataset as in max-information, perfect generalization, or replicability. Meanwhile, differential privacy, max-information, and perfect generalization quantify the sensitivity of the algorithm's output in a weaker way than replicability; the former three notions only require that the distributions on outputs are similar, whereas replicability demands that precisely the same output realization is obtained with high probability.

Nevertheless, the (surprising!) technical connections between these definitions have enabled substantial progress on the fundamental questions in their respective areas. For example, it was exactly the adaptive generalization guarantees of differential privacy that kickstarted the framework of adaptive data analysis from [15]; the definition of max-information was subsequently introduced [14] to unify existing analyses based on differential privacy and description length bounds. As another illustration, variants of replicability

[1] [23] stated this definition under the name "reproducibility." See Section 2.6 for a discussion of why we refer to it as "replicability" instead.

were introduced in [9, 19, 20] for purely technical reasons, as it was observed that such algorithms could be immediately used to construct differentially private ones. This connection was essential in proving the characterization of private PAC learnability in terms of the Littlestone dimension from online learning [2, 9]. In fact, this characterization shows, that, in principle a private PAC learner using n samples can be converted to a replicable PAC learner using a number of samples that is an exponential tower of height n, but it is non-constructive and does not suggest what such a learner looks like in general.

1.1 Our Main Results

1.1.1 Equivalences. Our main result is a complete characterization of the relationships between these quantities. We prove that all four central stability notions — replicability, differential privacy, perfect generalization, and bounded max-information w.r.t. product distributions — are equivalent to one another via constructive conversions that incur at most a near-quadratic overhead in sample complexity.

Our equivalences apply to an abstract and broad class of *statistical tasks* that capture learning from i.i.d. samples from a population. An instance of such a task is obtained by considering a distribution D from a pre-specified family of distributions. Given i.i.d. samples from D, the goal of a learning algorithm is to produce an outcome that is "good" for D with high probability. This formulation of a statistical task captures problems such as PAC learning, where a sample from D is a pair $(x, f(x)) \in \mathcal{X} \times \{0,1\}$ where x is drawn from an arbitrary marginal distribution over $\mathcal{X}$, and f is an arbitrary function from a fixed concept class H. A "good" outcome for such a distribution D is a hypothesis $h : \mathcal{X} \to \{0,1\}$ that well-approximates f on D. Many other objectives such as regression, distribution parameter estimation, distribution learning, hypothesis testing, and confidence interval construction can be naturally framed as statistical tasks. (See Section 6.4 for other examples.)

Figure 1 illustrates the known relationships between the various stability notions that hold with respect to any statistical task.

From these equivalences we obtain the following consequences, resolving several open questions.

Sample-efficient replicable algorithms. Any differentially private algorithm solving a statistical task (with a finite outcome space) can be converted into a replicable algorithm solving the same task with a near-quadratic blowup in its sample complexity. Thus, the wealth of research on private algorithm design can be brought to bear on designing replicable algorithms. We illustrate this algorithmic paradigm by describing new replicable algorithms for some PAC learning, distribution parameter estimation, and distribution testing problems in Section 6.4.

Equivalence between perfect generalization and differential privacy. For simplicity, the relationships summarized in Figure 1 are stated in terms of a one-way variant of perfect generalization, where only the inequality on the right of (1) is required to hold. But the original two-way definition turns out to be statistically equivalent for tasks with a finite outcome space. This is because a one-way perfectly generalizing algorithm can be converted to a replicable algorithm using Theorem 3.17, and Theorem 3.19 actually yields the stronger conversion back to a two-way perfectly generalizing algorithm (See Theorem 6.3). Thus, an (ε, δ)-differentially private algorithm (with a finite outcome space) can be converted to a perfectly generalizing one solving the same statistical task with a near quadratic blow-up in sample complexity. This resolves an open question of [12]. Their work also gave a conversion from perfectly generalizing algorithms to differentially private ones with no sample complexity overhead, and while their transformation preserves accuracy for (agnostic) PAC learning, it is not clear how to analyze it for general statistical tasks. Our conversion from perfect generalization to replicability and then to differential privacy holds for all statistical tasks with a finite outcome space.

Converting item-level to user-level privacy. Consider a "user-level" learning scenario in which n individuals each hold m training examples drawn i.i.d. from the same distribution. When is (ε, δ)-differentially private learning possible if we wish to guarantee privacy with respect to changing *all* of any individual's samples at once? Ghazi, Kumar, and Manurangsi [20] showed that this is possible when $n \geq O(\log(1/\delta)/\varepsilon)$ and the task admits a replicable learner. For the special case of PAC learning a concept class H, they argued that this implies a user-level private learning algorithm whenever H is privately PAC learnable with respect to changing a single *sample*. They posed the open problem of extending this result beyond PAC learning, e.g., to private regression [21, 25]. Our conversion from any differentially private algorithm to a replicable one implies that such a transformation is possible for *any* statistical task with a finite outcome space (Section 6.1). Moreover, one can always take each indvidual's number of samples m to be nearly quadratic in the sample complexity of the original item-level private learner.

Amplifying differential privacy parameters. While almost all (ε, δ)-differentially private algorithms enjoy a mild $\propto \log(1/\delta)$ dependence in their sample complexity on the parameter δ, it was not known how to achieve this universally, say by amplifying large values of δ to asymptotically smaller ones. [9] showed that for private PAC learning, such amplification is possible in principle, but posed the open question of giving an explicit amplification algorithm. By converting an (ε, δ)-differentially private algorithm with weak parameters to a replicable one, and then back to a differentially private one with strong parameters, we resolve this question for the general class of statistical tasks with a finite outcome space, and with a much milder sample complexity blowup (Section 6.2).

Agnostic-to-realizable reductions for distribution-family learning. [22] introduced a simple and flexible framework for converting realizable PAC learners to agnostic learners without relying on uniform convergence arguments. The framework applies to diverse settings such as robust learning, fair learning, partial learning, and (as observed in this work) replicable learning, with differential privacy providing a notable exception.[2] While an agnostic-to-realizable reduction for private PAC learning is known [1, 7], it relies on uniform convergence and is only known to hold in the distribution-free PAC model. By converting a realizable private learner to a realizable replicable learner, then to an agnostic replicable learner, and back

[2] We note the technique we introduce to adapt [22] to the replicable setting has no clear translation to the private setting.

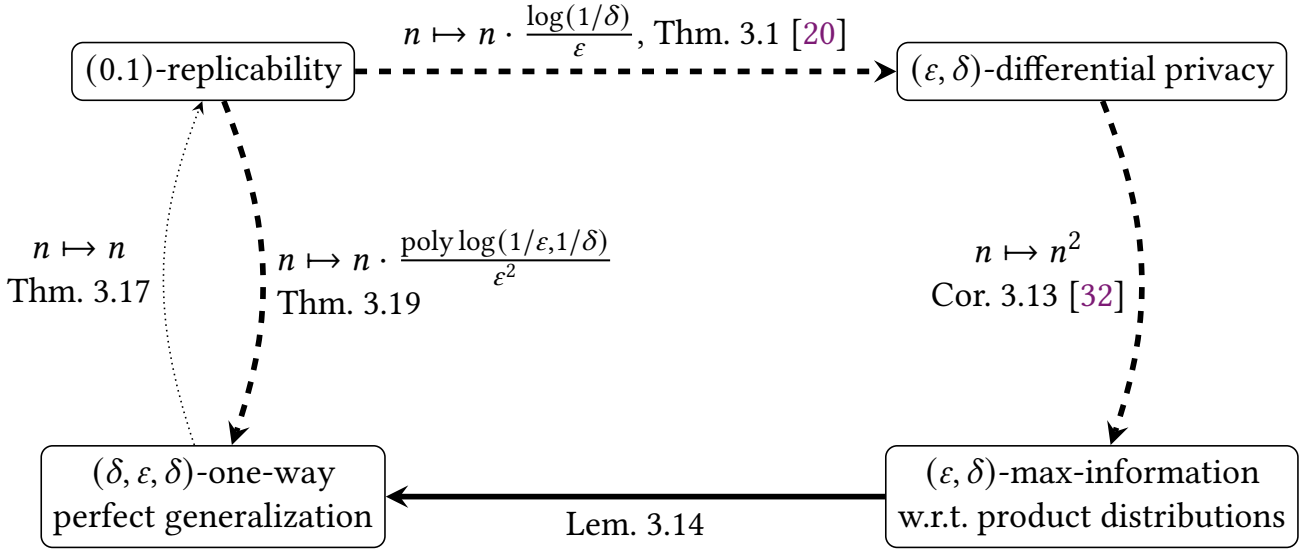

Figure 1: The solid arrow from A to B means that every algorithm satisfying A also satisfies B. A dashed arrow means that for every statistical task, a solution satisfying A can be computationally efficiently transformed into a solution satisfying B with the stated blowup in sample complexity. The thin dotted arrow means an explicit transformation exists, but is not always computationally efficient, and assumes the outcome space is finite.
This figure suppresses constant factors everywhere and polynomial factors in δ, assumes ε is below a sufficiently small constant, and assumes that δ is a sufficiently small inverse polynomial in n.

to an agnostic private learner, we obtain a reduction that works in the absence of uniform convergence (Section 6.3). In particular, this reduction applies to the *distribution-family* learning model, where one is promised that the marginal distribution on unlabeled examples comes from a pre-specified family of distributions.

1.1.2 Separating Stability: Computational Barriers and the Complexity of Correlated Sampling. All of the transformations appearing in Figure 1 preserve computational efficiency, with the lone exception of the transformation from perfectly generalizing algorithms to replicable ones. This transformation makes use of the technique of *correlated sampling* from the distribution of outputs of a perfectly generalizing algorithm A when run on a fixed sample S (elaborated on more in Sections 1.2 and 2.5). This step can be explicitly implemented via rejection sampling from the output space of A, with the rejection threshold determined by the probability mass function of $A(S)$, but in general it is not computationally efficient.

We show that under cryptographic assumptions, this is inherent (Section 4). Specifically, we show that under standard assumptions in public-key cryptography, there exists a statistical task that admits an efficient differentially private algorithm, but does not have any efficient replicable algorithm. The task is defined in terms of a public-key encryption scheme with the following rerandomizability property: Given a ciphertext $\text{Enc}(\mathbf{pk}, b)$, there is an efficient algorithm producing a uniformly random encryption of b. Fixing such a rerandomizable PKE, the statistical task is as follows. Given a dataset consisting of random encryptions of the form $\text{Enc}(\mathbf{pk}, b)$ where $\mathbf{pk}$ is a fixed public key and $b \in \{0, 1\}$ is a fixed bit, output any encryption of b.

One can solve this problem differentially privately, essentially by choosing a random ciphertext from the input dataset and rerandomizing it. On the other hand, there is no efficient replicable algorithm for this task. If there were, then one could use the public key to produce many encryptions of 0 and 1 and run the replicable algorithm on the results to produce canonical ciphertexts c_0 and c_1, respectively. Then, given an unknown ciphertext, one could repeatedly rerandomize it, run the replicable algorithm on the results, and compare the answer to c_0 and to c_1 to identify the underlying plaintext.

We also show that cryptographic assumptions are necessary even to separate replicability from perfect generalization. Recalling again that the bottleneck in computationally equating the two notions is in implementing correlated sampling, we show in Section 4.2 that if one-way functions do not exist, then correlated sampling is always tractable. In addition to addressing a natural question about the complexity of correlated sampling, this shows that function inversion enables an efficient transformation from perfectly generalizing algorithms into replicable ones. (See Section 2.5 for more discussion.)

1.1.3 Separating Stability: Statistical Barriers. Our equivalences show that the sample complexities of perfectly generalizing and replicable learning are essentially equivalent. Moreover: (1) An approximate-DP algorithm can be converted to a perfectly generalizing/replicable algorithm with near-quadratic blowup; and (2) A perfectly generalizing/replicable algorithm can be converted to an approximate-DP one using roughly the same number of samples. We prove that both of these conversions are optimal by showing:

(1) **Quadratic separations between differential privacy and perfect generalization/replicability.** We first consider the problem of estimating the parameters of a product of d Bernoulli distributions. By simply taking the empirical mean of an input dataset, this problem can be solved using $O(\log d)$ without any stability constraints. However, with differential privacy, it is known that $\tilde{\Theta}(\sqrt{d})$ samples are necessary and sufficient. By adapting the "fingerprinting" method underlying these privacy lower bounds [10, 11, 18] to perfect generalization, we prove that any perfectly generalizing or replicable algorithm for this problem requires $\tilde{\Omega}(d)$ samples (Section 5.1).
By reducing from a variant of this one-way marginals problem, we also show a general lower bound for replicable agnostic learning. Namely, we show that every concept class H requires $\tilde{\Omega}(VC(H)^2)$ samples. For concept classes of maximal VC dimension $VC(H) = \log|H|$, this too gives a quadratic separation between replicable learning and both private and unconstrained learning (Section 5.2).
(2) **No separation between differential privacy and perfect generalization/replicability.** Complementing our lower bounds, we also show that every finite class H can be replicably PAC learned (in the realizable setting) to error α with sample complexity $\tilde{O}_H(1/\alpha)$ (Section 5.3). Up to logarithmic factors, this matches the learning rate achievable for both unconstrained and differentially private learning. Our learner works by selecting a random threshold v, and selecting a random concept from H whose error with respect to the sample is at most v. A more involved random thresholding strategy also yields an agnostic learner with sample complexity $\tilde{O}_H(1/\alpha^2)$.

1.2 Overview of Proofs of Equivalences

Perfect generalization is equivalent to replicability. Recall that an algorithm is replicable if it is likely to produce exactly the same output when run on two independent samples from any given population. Replicability appears to be a dramatic strengthening of perfect generalization, which only requires the distributions of $A(S)$ and $A(S')$ to be statistically close. Nevertheless, we prove that perfectly generalizing algorithms can always be converted to replicable ones whenever the output space $\mathcal{Y}$ is finite (Theorem 3.17). This can be done via a primitive called *correlated sampling* (See Section 2.5). A correlated sampling algorithm for a class of distributions $\mathcal{P} = \{P\}$ is a procedure $CS(P, r)$ such that 1) $CS(P, r)$ produces a sample distributed according to P when provided a uniformly random input r, and 2) Whenever $P, Q \in \mathcal{P}$ satisfy $d_{\mathrm{TV}}(P, Q) \leq \eta$, we have $\mathbf{Pr}[CS(P, r) = CS(Q, r)] \geq 1 - O(\eta)$. That is, applying correlated sampling to two similar distributions results in the same output with high probability – exactly what is needed for replicability. We actually prove a stronger theorem, showing that the larger class of *one-way* perfectly generalizing algorithms (where only the right-hand inequality in 1 holds) are replicable via correlated sampling.

Conversely, we show how to convert replicable algorithms to perfectly generalizing ones (Theorem 3.19). While a ρ-replicable algorithm is automatically also a $(\beta = O(\rho), \varepsilon = 0, \delta = O(\rho))$-perfectly generalizing one, these parameters are too weak for applications where one wants to take β, δ to be inverse polynomial in the dataset size n (e.g., to prove the lower bounds in Section 5). To obtain a perfectly generalizing algorithm with stronger parameters, we repeatedly run the replicable algorithm using $k = O(\log(1/\delta))$ different sequences of coin tosses $r_1, \ldots, r_k$, and using $\tilde{O}(1/\varepsilon^2)$ independent samples for each sequence of coin tosses. Using the exponential mechanism from differential privacy [28], we select an outcome y_i that appears approximately the most frequently amongst these repetitions in a manner that ensures $(\beta = \delta, \varepsilon, \delta)$-perfect generalization. This strategy allows us to obtain inverse polynomial β, δ parameters with only a logarithmic multiplicative overhead in the number of samples.

Bounded max-information implies perfect generalization. In Lemma 3.14, we show that bounded max-information implies one-way perfect generalization with similar parameters. Namely, if an algorithm A has (ε, δ)-max-information with respect to product distributions, then it is also $(\sqrt{\delta}, 2\varepsilon, \sqrt{\delta})$-one-way perfectly generalizing. The idea is to take the simulator distribution Sim_D to be the distribution of $A(S')$, where the randomness is taken over both the coin tosses of A *and* the randomness of a sample $S' \sim D$. A similar argument is implicit in [4, Proof of Lemma 4.5]. Then by combining Theorems 3.17 and 3.19, it follows that bounded max-information also implies perfect generalization for finite outcome spaces (Theorem 6.3).

Replicability implies differential privacy. In Theorem 3.1 we show that replicability implies differential privacy. Given a replicable algorithm, one can run it $k = O(\log(1/\delta)/\varepsilon)$ times using the same sequence of coin tosses, but on independent samples, producing outcomes $y_1, \ldots, y_k$. Replicability ensures that most of these outcomes are the same with high probability, and so this common outcome can be selected in a standard differentially private way. This argument appears in the differential privacy literature as a conversion from "globally stable" and "pseudo-globally stable" learners to private ones [9, 19, 20]. Our presentation of Theorem 3.1 includes an additional amplification step that avoids union bounding over correctness, making the conversion suitable for a broader range of parameters.

Differential privacy implies bounded max-information. The final conversion in Figure 1 is from differentially private algorithms to algorithms with bounded max-information. This argument is implicit in [32] and we show how it follows from their work here (Corollaries 3.12 and 3.13).

1.3 Further Discussion of Related Work

Several elements of our approach were inspired by Ghazi, Kumar, and Manurangsi's study of the relationship between user-level and item-level differentially private learning [20]. They introduced a notion of "pseudo-global stability" that is essentially the same as replicability, and showed that it implies differential privacy. Correlated sampling also played a crucial role in their work by allowing individuals to use shared randomness to reach consensus on a learned hypothesis. In fact, it provided a key step in their conversion from "list globally stable" algorithms [19] (learning algorithms

that output a short list of hypotheses, one of which is almost guaranteed to be canonical for the given distribution) to pseudo-globally stable ones.

Stability in learning has a long history as a tool for ensuring generalization. Early work [8, 13, 33, 35] showed that the stability of a learning algorithm with respect to a specific loss function could ensure strong generalization guarantees with respect to that loss. A more recent literature has focused on stability notions that are not tied to a specific loss, and which ideally are robust under post-processing and adaptive composition. This includes understanding the generalization guarantees of differential privacy [6, 14, 15, 24, 31, 32, 38] and other constraints on the information-theoretic relationship between the input and output of a learning algorithm [5, 27, 30, 34, 36, 37]. A related line of work considers more "semantic" notions of stability, defining it in terms of the difficulty of inferring properties specific to the sample rather than of the underlying distribution [4, 12, 29]. Perfect generalization, one of the main definitions we study in this work, was introduced by [12] and is a special case of *typical stability* that was introduced in independent work of Bassily and Freund [4].

Independent of this work, [26] study similar relationships between notions of stability. They focus on the PAC-learning setting, where they show a statistical equivalence between differential privacy, replicability, and a notion called "TV-indistinguishability" which can be thought of as a special case of perfect generalization with $\varepsilon = 0$. To clarify the differences between our work and [26], first recall how we obtain replicability from differential privacy:

- First, we exploit existing connections between privacy and bounded max-information from [32] to obtain an algorithm with bounded max-information from a differentially private one.
- We prove that bounded max-information implies perfect generalization.
- We then show that we can obtain a replicable algorithm from a perfectly generalizing one by applying correlated sampling to its distribution over outputs. The relevant output distribution is induced by fixing an input sample of the perfectly generalizing algorithm and redrawing its internal randomness.

Recall that the correlated sampling procedure may not be efficient, and that we assume the output domain of the differentially private algorithm is finite.

The work of [26] follows a different approach. First, they start from a differentially private PAC learner, rather than a differentially private algorithm for a general statistical task, and factor through TV-indistinguishability and Littlestone dimension. More specifically:

- They first observe a similar equivalence of replicability and TV-indistinguishability for general statistical tasks.
- They then show that a private PAC learner implies the existence of a TV-indistinguishable learner, leveraging results from [2] showing that private PAC learning implies finite Littlestone dimension, and results from [19, 20] showing that finite Littestone dimension implies list global-stability.

Our approach gives us a constructive procedure for converting a private algorithm for a general statistical task into a replicable algorithm, so long as the private algorithm has finite range. Our transformations induce a modest sample complexity increase, resulting in a replicable algorithm with sample complexity n^2, given a private learner with sample complexity n. By contrast, the results of [26], while non-constructive, apply to countably infinite domains (and therefore to some uncountably infinite ranges). However, their results go through Littlestone dimension, which may be an exponential tower in n, and so they obtain sample complexity bounds which are an exponential tower in n as well.

1.4 Open Problems

We highlight several directions and open problems for future work.

(1) Is a transformation from (one-way) perfectly generalizing algorithms to replicable algorithms possible for infinite output spaces in general? While correlated sampling introduces no sample complexity overhead in terms of the output space, it is only known to be possible when the output space is finite or the class of distributions to be sampled from is structured. (E.g., the distributions in the class all have uniformly bounded Radon-Nikodym derivative with respect to some fixed base measure).[3] In independent work, [26] make progress towards this goal by giving a transformation from TV-indistinguishability to replicability when there are only countably many options for the TV-indistinguishable algorithm $\{A(S)\}_{S\in X^n}$. In the full version, we show that $(\beta, \varepsilon, \delta)$-one-way perfect generalization implies $(4\varepsilon+2\delta+2\beta)$-TV indistinguishability, and so the result of [26] gives the following corollary.

   Corollary 1.4. *Fix $n \in \mathbb{N}$, $\beta, \varepsilon, \delta \in (0, 1]$. Let $\mathcal{X}$ be a countable domain and $A : \mathcal{X}^n \to \mathcal{Y}$ be a $(\beta, \varepsilon, \delta)$-one-way perfectly generalizing algorithm for a statistical task. Then there exists an algorithm $A' : \mathcal{X}^n \to \mathcal{Y}$ that is $\left(\frac{2\rho}{1+\rho}\right)$-replicable for $\rho = 4\varepsilon + 2\delta + 2\beta$, and for all $S \in \mathcal{X}^n$, $A(S) = A'(S)$.*

   Whether a transformation exists for general measure spaces remains open.[4] In the full version we discuss the *list heavy-hitters* problem that may be a candidate for separating perfect generalization from replicability over infinite output spaces.

(2) What are the minimal cryptographic assumptions under which a computational separation between replicability and differential privacy exists? Our results show that one-way functions are necessary, while public-key assumptions are sufficient.

(3) [23, Lemma A.7] showed that replicable algorithms compose adaptively. That is, a sequence of k adaptively chosen ρ-replicable algorithms yields a transcript that is $O(k\rho)$-replicable. One way to interpret this result is as follows: Given a sequence of k analyses that are each (0.01)-replicable using a sample of size n, one can amplify their individual

[3] Formally, such a case would fall into a restricted notion of correlated sampling over a subset of distributions, similar to the multiple coupling of [3].

[4] We note that in the PAC-setting one can resolve this issue via factoring through Littlestone Dimension and [23]'s heavy-hitters, but this results in tower sample complexity.

replicability parameters to $O(1/k)$ at the expense of increasing their sample complexity to $O(k^2 n)$. This yields a (0.01)-replicable algorithm for performing all k analyses at a sample cost of $O(k^2 n)$.

Our conversions between replicability and differential privacy yield a different tradeoff, at least for simulating non-adaptive composition. Given k analyses that are each (0.01)-replicable using a sample of size n, one can convert them to $\tilde{O}(1/\sqrt{k})$-differentially private algorithms each using a sample of size $\tilde{O}(\sqrt{k}n)$. "Advanced" composition of differential privacy [17] yields an $(0.01, \delta)$-differentially private algorithm using $\tilde{O}(\sqrt{k}n)$ samples, which can then be turned back into a (0.01)-replicable algorithm using $\tilde{O}(kn^2)$ samples.

What is the optimal sample cost for conducting, or at least statistically simulating, the (adaptive) composition of k replicable algorithms? Is it possible to do so at a cost of $O(kn)$ samples?

(4) In the full version, we give a direct replicable algorithm for the task of realizable PAC learning of finite classes with sample cost inverse linear in the accuracy parameter α. (As opposed to inverse quadratic, which is what applying the reduction from replicability to approximate DP gives – see Theorem 6.13 and the following discussion.) Are there other natural problems for which there are (perhaps more dramatic) separations between what's achievable via directly constructing a replicable algorithm for a task, and what's achievable using our reduction to approximate DP? For example, can discrete distributions over $[k]$ be replicably estimated using $O(k)$ samples (as opposed to quadratic in k, which is what is obtained through our reduction)? Can the mean of a d-variate Gaussian with unknown covariance be estimated directly using $O(d)$ samples (as opposed to quadratic in d, which is what is obtained through our reduction)? Even more ambitiously, is it possible to characterize the types of problems for which our reduction from replicability to approximate DP gives tight bounds?

(5) To what extent is replicability preserved under distributional shift? In the full version, we give a simple argument showing that a ρ-replicable algorithm is $\rho(1-\delta)^{2m}$-replicable across two close distributions. Are there tighter replicability and non-replicability bounds for specific families of distributions, problems, and algorithms under distributional shifts?

ACKNOWLEDGMENTS

We thank Adam Smith for helpful discussions on max information, Zhiwei Wu for pointing us to perfect generalization, Shay Moran for helpful discussions regarding the relation of our work to [26], and Christopher Ye for helpful comments on a prior version of this manuscript. The views expressed in this paper are those of the authors and not those of the U.S. Census Bureau or any other sponsor.

Supported by NSF Awards CCF-1947889, CNS-2046425, CNS-2040249, DGE-1650112, AF: Medium 2212136, DGE-1650112. Supported by a Sloan Research Fellowship, the Simons Foundation Collaboration on the Theory of Algorithmic Fairness, and Cooperative Agreement CB20ADR0160001 with the Census Bureau.

REFERENCES

[1] Noga Alon, Amos Beimel, Shay Moran, and Uri Stemmer. 2020. Closure properties for private classification and online prediction. In *Conference on Learning Theory*. PMLR, 119–152.

[2] Noga Alon, Roi Livni, Maryanthe Malliaris, and Shay Moran. 2019. Private PAC learning implies finite Littlestone dimension. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 852–860.

[3] Omer Angel and Yinon Spinka. 2019. Pairwise optimal coupling of multiple random variables. https://doi.org/10.48550/ARXIV.1903.00632

[4] Raef Bassily and Yoav Freund. 2016. Typicality-Based Stability and Privacy. *CoRR* abs/1604.03336 (2016). arXiv:1604.03336 http://arxiv.org/abs/1604.03336

[5] Raef Bassily, Shay Moran, Ido Nachum, Jonathan Shafer, and Amir Yehudayoff. 2018. Learners that Use Little Information. In *Algorithmic Learning Theory, ALT 2018, 7-9 April 2018, Lanzarote, Canary Islands, Spain (Proceedings of Machine Learning Research, Vol. 83)*, Firdaus Janoos, Mehryar Mohri, and Karthik Sridharan (Eds.). PMLR, 25–55. http://proceedings.mlr.press/v83/bassily18a.html

[6] Raef Bassily, Kobbi Nissim, Adam D. Smith, Thomas Steinke, Uri Stemmer, and Jonathan R. Ullman. 2016. Algorithmic stability for adaptive data analysis. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, Daniel Wichs and Yishay Mansour (Eds.). ACM, 1046–1059. https://doi.org/10.1145/2897518.2897566

[7] Amos Beimel, Kobbi Nissim, and Uri Stemmer. 2016. Private Learning and Sanitization: Pure vs. Approximate Differential Privacy. *Theory Comput.* 12, 1 (2016), 1–61. https://doi.org/10.4086/toc.2016.v012a001

[8] Olivier Bousquet and André Elisseeff. 2002. Stability and generalization. *The Journal of Machine Learning Research* 2 (2002), 499–526.

[9] Mark Bun, Roi Livni, and Shay Moran. 2020. An equivalence between private classification and online prediction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 389–402.

[10] Mark Bun, Thomas Steinke, and Jonathan R. Ullman. 2019. Make Up Your Mind: The Price of Online Queries in Differential Privacy. *J. Priv. Confidentiality* 9, 1 (2019). https://doi.org/10.29012/jpc.655

[11] Mark Bun, Jonathan R. Ullman, and Salil P. Vadhan. 2018. Fingerprinting Codes and the Price of Approximate Differential Privacy. *SIAM J. Comput.* 47, 5 (2018), 1888–1938. https://doi.org/10.1137/15M1033587

[12] Rachel Cummings, Katrina Ligett, Kobbi Nissim, Aaron Roth, and Zhiwei Steven Wu. 2016. Adaptive Learning with Robust Generalization Guarantees. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016 (JMLR Workshop and Conference Proceedings, Vol. 49)*, Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir (Eds.). JMLR.org, 772–814. http://proceedings.mlr.press/v49/cummings16.html

[13] L. Devroye and T. Wagner. 1979. Distribution-free performance bounds for potential function rules. *IEEE Transactions on Information Theory* 25, 5 (1979), 601–604. https://doi.org/10.1109/TIT.1979.1056087

[14] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. 2015. Generalization in Adaptive Data Analysis and Holdout Reuse. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett (Eds.). 2350–2358. https://proceedings.neurips.cc/paper/2015/hash/bad5f33780c42f2588878a9d07405083-Abstract.html

[15] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. 2015. Preserving Statistical Validity in Adaptive Data Analysis. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, Rocco A. Servedio and Ronitt Rubinfeld (Eds.). ACM, 117–126. https://doi.org/10.1145/2746539.2746580

[16] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2016. Calibrating Noise to Sensitivity in Private Data Analysis. *J. Priv. Confidentiality* 7, 3 (2016), 17–51. https://doi.org/10.29012/jpc.v7i3.405

[17] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. 2010. Boosting and Differential Privacy. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*. IEEE Computer Society, 51–60. https://doi.org/10.1109/FOCS.2010.12

[18] Cynthia Dwork, Adam D. Smith, Thomas Steinke, Jonathan R. Ullman, and Salil P. Vadhan. 2015. Robust Traceability from Trace Amounts. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, Venkatesan Guruswami (Ed.). IEEE Computer Society, 650–669. https://doi.org/10.1109/FOCS.2015.46

[19] Badih Ghazi, Noah Golowich, Ravi Kumar, and Pasin Manurangsi. 2021. Sample-efficient proper PAC learning with approximate differential privacy. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 183–196.

[20] Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. 2021. User-Level Differentially Private Learning via Correlated Sampling. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, Marc'Aurelio

Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (Eds.). 20172–20184. https://proceedings.neurips.cc/paper/2021/hash/a89cf525e1d9f04d16ce31165e139a4b-Abstract.html

[21] Noah Golowich. 2021. Differentially Private Nonparametric Regression Under a Growth Condition. In *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA (Proceedings of Machine Learning Research, Vol. 134)*, Mikhail Belkin and Samory Kpotufe (Eds.). PMLR, 2149–2192. http://proceedings.mlr.press/v134/golowich21a.html

[22] Max Hopkins, Daniel M. Kane, Shachar Lovett, and Gaurav Mahajan. 2022. Realizable Learning is All You Need. In *Conference on Learning Theory, 2-5 July 2022, London, UK (Proceedings of Machine Learning Research, Vol. 178)*, Po-Ling Loh and Maxim Raginsky (Eds.). PMLR, 3015–3069. https://proceedings.mlr.press/v178/hopkins22a.html

[23] Russell Impagliazzo, Rex Lei, Toniann Pitassi, and Jessica Sorrell. 2022. Reproducibility in learning. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, Stefano Leonardi and Anupam Gupta (Eds.). ACM, 818–831. https://doi.org/10.1145/3519935.3519973

[24] Christopher Jung, Katrina Ligett, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi, and Moshe Shenfeld. 2020. A New Analysis of Differential Privacy's Generalization Guarantees. In *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, January 12-14, 2020, Seattle, Washington, USA (LIPIcs, Vol. 151)*, Thomas Vidick (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 31:1–31:17. https://doi.org/10.4230/LIPIcs.ITCS.2020.31

[25] Young Hun Jung, Baekjin Kim, and Ambuj Tewari. 2020. On the Equivalence between Online and Private Learnability beyond Binary Classification. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS'20)*. Curran Associates Inc., Red Hook, NY, USA, Article 1401, 10 pages.

[26] Alkis Kalavasis, Amin Karbasi, Shay Moran, and Grigoris Velegkas. 2023. Statistical Indistinguishability of Learning Algorithms. Personal communication.

[27] Katrina Ligett and Moshe Shenfeld. 2019. A Necessary and Sufficient Stability Notion for Adaptive Generalization. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (Eds.). 11481–11490. https://proceedings.neurips.cc/paper/2019/hash/c5df4f4eabf1cbcfeb50fbbf97c5289f-Abstract.html

[28] Frank McSherry and Kunal Talwar. 2007. Mechanism Design via Differential Privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS '07)*. IEEE Computer Society, USA, 94–103. https://doi.org/10.1109/FOCS.2007.41

[29] Kobbi Nissim, Adam Smith, Uri Stemmer, Thomas Steinke, and Jonathan Ullman. 2018. The limits of post-selection generalization. *Advances in Neural Information Processing Systems* 31 (2018).

[30] Maxim Raginsky, Alexander Rakhlin, Matthew Tsao, Yihong Wu, and Aolin Xu. 2016. Information-theoretic analysis of stability and bias of learning algorithms. In *2016 IEEE Information Theory Workshop, ITW 2016, Cambridge, United Kingdom, September 11-14, 2016*. IEEE, 26–30. https://doi.org/10.1109/ITW.2016.7606789

[31] Ryan Rogers, Aaron Roth, Adam D. Smith, Nathan Srebro, Om Thakkar, and Blake E. Woodworth. 2020. Guaranteed Validity for Empirical Approaches to Adaptive Data Analysis. In *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26-28 August 2020, Online [Palermo, Sicily, Italy] (Proceedings of Machine Learning Research, Vol. 108)*, Silvia Chiappa and Roberto Calandra (Eds.). PMLR, 2830–2840. http://proceedings.mlr.press/v108/rogers20a.html

[32] Ryan M. Rogers, Aaron Roth, Adam D. Smith, and Om Thakkar. 2016. Max-Information, Differential Privacy, and Post-selection Hypothesis Testing. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, Irit Dinur (Ed.). IEEE Computer Society, 487–494. https://doi.org/10.1109/FOCS.2016.59

[33] W. H. Rogers and T. J. Wagner. 1978. A Finite Sample Distribution-Free Performance Bound for Local Discrimination Rules. *The Annals of Statistics* 6, 3 (1978), 506 – 514. https://doi.org/10.1214/aos/1176344196

[34] Daniel Russo and James Zou. 2016. Controlling Bias in Adaptive Data Analysis Using Information Theory. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, Cadiz, Spain, May 9-11, 2016 (JMLR Workshop and Conference Proceedings, Vol. 51)*, Arthur Gretton and Christian C. Robert (Eds.). JMLR.org, 1232–1240. http://proceedings.mlr.press/v51/russo16.html

[35] Shai Shalev-Shwartz, Ohad Shamir, Nathan Srebro, and Karthik Sridharan. 2010. Learnability, stability and uniform convergence. *The Journal of Machine Learning Research* 11 (2010), 2635–2670.

[36] Thomas Steinke and Lydia Zakynthinou. 2020. Reasoning About Generalization via Conditional Mutual Information. In *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria] (Proceedings of Machine Learning Research, Vol. 125)*, Jacob D. Abernethy and Shivani Agarwal (Eds.). PMLR, 3437–3452. http://proceedings.mlr.press/v125/steinke20a.html

[37] Aolin Xu and Maxim Raginsky. 2017. Information-theoretic analysis of generalization capability of learning algorithms. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 2524–2533. https://proceedings.neurips.cc/paper/2017/hash/ad71c82b22f4f65b9398f76d8be4c615-Abstract.html

[38] Tijana Zrnic and Moritz Hardt. 2019. Natural Analysts in Adaptive Data Analysis. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 7703–7711. http://proceedings.mlr.press/v97/zrnic19a.html

Received 2022-11-07; accepted 2023-02-06