Partially Symmetric Functions are Efficiently Isomorphism-Testable

Eric Blais
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA.
Email: eblais@cs.cmu.edu

Amit Weinstein
Blavatnik School of Computer Science
Tel Aviv University
Tel Aviv, Israel.
Email: amitw@tau.ac.il

Yuichi Yoshida
National Institute of Informatics, and
Preferred Infrastructure, Inc.
Tokyo, Japan.
Email: yyoshida@nii.ac.jp

***Abstract*—Given a Boolean function f, the *f-isomorphism testing* problem requires a randomized algorithm to distinguish functions that are identical to f up to relabeling of the input variables from functions that are far from being so. An important open question in property testing is to determine for which functions f we can test f-isomorphism with a constant number of queries. Despite much recent attention to this question, essentially only two classes of functions were known to be efficiently isomorphism testable: symmetric functions and juntas.**

We unify and extend these results by showing that all *partially symmetric* functions—functions invariant to the reordering of all but a constant number of their variables—are efficiently isomorphism-testable. This class of functions, first introduced by Shannon, includes symmetric functions, juntas, and many other functions as well. We conjecture that these functions are essentially the only functions efficiently isomorphism-testable.

To prove our main result, we also show that partial symmetry is efficiently testable. In turn, to prove this result we had to revisit the junta testing problem. We provide a new proof of correctness of the nearly-optimal junta tester. Our new proof replaces the Fourier machinery of the original proof with a purely combinatorial argument that exploits the connection between sets of variables with low influence and intersecting families.

Another important ingredient in our proofs is a new notion of *symmetric influence*. We use this measure of influence to prove that partial symmetry is efficiently testable and also to construct an efficient sample extractor for partially symmetric functions. We then combine the sample extractor with the testing-by-implicit-learning approach to complete the proof that partially symmetric functions are efficiently isomorphism-testable.

***Keywords*-Boolean functions; property testing; partial symmetry;**

I. Introduction

Property testing considers the following general problem: given a property $\mathcal{P}$, identify the minimum number of queries required to determine with high probability whether an input has the property $\mathcal{P}$ or whether it is far from $\mathcal{P}$. This question was first formalized by Rubinfeld and Sudan [1].

Definition 1 (Property tester)**.** Let $\mathcal{P}$ be a set of Boolean functions. An *ϵ-tester* for $\mathcal{P}$ is a randomized algorithm which queries an unknown function $f : \{0,1\}^n \to \{0,1\}$ on a small number of inputs and

1) Accepts with probability at least $2/3$ when $f \in \mathcal{P}$;
2) Rejects with probability at least $2/3$ when f is ϵ-far from $\mathcal{P}$,

where f is *ϵ-far* from $\mathcal{P}$ if $\mathrm{dist}(f,g) := |\{x \in \{0,1\}^n \mid f(x) \neq g(x)\}| \geq \epsilon 2^n$ holds for every $g \in \mathcal{P}$.

Goldreich, Goldwasser, and Ron [2] extended the scope of this definition to graphs and other combinatorial objects. Since then, the field of property testing has been very active. For an overview of recent developments, we refer the reader to the surveys [3], [4] and the book [5].

A notable achievement in the field of property testing is the complete characterization of graph properties that are testable with a constant number of queries [6]. An ambitious open problem is obtaining a similar characterization for properties of Boolean functions. Recently there has been a lot of progress on the restriction of this question to properties that are closed under linear or affine transformations [7], [8]. More generally, one might hope to settle this open problem for all properties of Boolean functions that are closed under relabeling of the input variables.

An important sub-problem of this open question is function isomorphism testing. Given a Boolean function f, the *f-isomorphism testing* problem is to determine whether a function g is isomorphic to f—that is, whether it is the same up to relabeling of the input variables—or *far* from being so. A natural goal, and the focus of this paper, is to characterize the set of functions for which isomorphism testing can be done with a constant number of queries.

A. Previous work

The function isomorphism testing problem was first raised by Fischer et al. [9]. They observed that fully symmetric functions are trivially isomorphism testable with a constant number of queries. They also showed that every *k-junta*, that is every function which depends on at most k of the input variables, is isomorphism testable with $\mathrm{poly}(k)$ queries. This bound was recently improved by Chakraborty et al. [10], who showed that $O(k \log k)$ queries suffice. These results imply that juntas on a constant number of variables are isomorphism testable with a constant number of queries.

The first lower bound for isomorphism testing was also provided by Fischer et al. [9]. They showed that for small

0272-5428/12 $26.00 © 2012 IEEE
DOI 10.1109/FOCS.2012.53

enough values of k, testing isomorphism to a k-linear function (i.e., a function that returns the parity of k variables) requires $\Omega(\log k)$ queries.[1] Following a series of recent works [11], [12], [13], the exact query complexity for testing isomorphism to k-linear functions has been determined to be $\tilde{\Theta}(\min(k, n-k))$.

More general lower bounds for isomorphism testing were obtained by Blais and O'Donnell [14]. In particular, they showed that testing isomorphism to *any* k-junta that is far from being a $(k-1)$-junta requires $\Omega(\log\log k)$ queries. This lower bound gives a large family of functions for which testing isomorphism requires a super-constant number of queries. Alon et al. proved even more general lower bounds showing that for almost every function f, testing isomorphism to f requires $\tilde{\Theta}(n)$ queries [15] (see also [16], [10]).

B. *Partially symmetric functions*

As seen above, the only functions which we know are isomorphism testable with a constant number of queries are fully symmetric functions and juntas. Our motivation for the current work was to see if we can unify and generalize the results to encompass a larger class of functions. While symmetric functions and juntas may seem unrelated, there is in fact a strong connection. Symmetric functions, of course, are invariant under any relabeling of the input variables. Juntas satisfy a similar but slightly weaker invariance property. For every k-junta, there is a set of at least $n-k$ variables such that the function is invariant to any relabeling of these variables. Functions that satisfy this condition are called *partially symmetric*.

Definition 2 (Partially symmetric functions)**.** For a subset $J \subseteq [n] := \{1,\dots,n\}$, a function $f : \{0,1\}^n \to \{0,1\}$ is *J-symmetric* if permuting the labels of the variables of J does not change the function. Moreover, f is called *t-symmetric* if there exists $J \subseteq [n]$ of size at least t such that f is J-symmetric.

Shannon first introduced partially symmetric functions as part of his investigation on the circuit complexity of Boolean functions [17]. He showed that while most functions require an exponential number of gates to compute, every partially symmetric function can be implemented much more efficiently. Research on the connection between partial symmetry and the complexity of Boolean functions has remained active ever since [18], [19], [20], [21], [22], [23], [24], [25], [26], [27].[2] Our results suggest that studying partially symmetric functions may also yield greater understanding of property testing on Boolean functions.

[1]More precisely, they showed that non-adaptive testers require $\tilde{\Omega}(\sqrt{k})$ queries. Here and in the rest of this section, tilde notation is used to hide logarithmic factors.

[2]Different definitions of partial symmetry have been introduced since the original work of Shannon [17]. All of these definitions are related and, in fact, many of them are equivalent [28].

C. *Our results*

The set of partially symmetric functions includes both juntas and symmetric functions, but the set also contains many other functions as well. A natural question is whether this entire class of functions is isomorphism testable with a constant number of queries. Our first main result gives an affirmative answer to this question.

Theorem 1. *For every $(n-k)$-symmetric function $f : \{0,1\}^n \to \{0,1\}$ there exists an ϵ-tester for f-isomorphism that performs $O(k \log k/\epsilon^2)$ queries.*

A simple modification of an argument in Alon et al. [15] can be used to show that the bound in the above theorem is tight up to logarithmic factors. Indeed by this argument, testing isomorphism to almost every $(n-k)$-symmetric function requires $\Omega(k)$ queries.

We believe that the theorem might also be best possible in a different way. That is, we conjecture that the set of partially symmetric functions is essentially the set of functions for which testing isomorphism can be done with a constant number of queries. We discuss this conjecture with some supporting evidence in Section VI.

The proof of our first main theorem follows the general outline of the proof that isomorphism testing to juntas can be done in a constant number of queries. The observation which allows us to make this connection is the fact that partially symmetric functions can be viewed as junta-like functions. More precisely, an $(n-k)$-symmetric function is a function that has k special variables where for each assignment for these variables, the restricted function is fully symmetric on the remaining $n-k$ variables.

The proof for testing isomorphism of juntas has two main components. The first is an efficient junta testing algorithm. This enables us to reject functions that are far from being juntas. The second is a query efficient sampler of the "core" of the input function given that the function is close to a junta. The sampler can then be used in order to verify if the two juntas are indeed isomorphic. We generalize both of these components for partially symmetric functions.

Our second main result, and the first component of the isomorphism tester, is an efficient algorithm for testing partial symmetry.

Theorem 2. *The property of being $(n-k)$-symmetric for $k < n/10$ is testable with $O(\frac{k}{\epsilon}\log\frac{k}{\epsilon})$ queries.*

The natural approach for proving this theorem is to try to generalize the result on junta testing in [29]. That result heavily relied on the notion of influence of variables. The *influence* of a set S of variables in a function f is the probability that $f(x) \neq f(y)$ when x is chosen uniformly at random and y is obtained from x by re-randomizing the values of x_i for each $i \in S$. The notion of influence characterizes juntas: when f is a k-junta, there is a set of size $n-k$ whose influence is 0, whereas when f is ϵ-far from being a k-junta, every set of size $n-k$ has influence at least ϵ.

We introduce a different notion of influence which we call *symmetric influence*. The symmetric influence of a set S of variables in f is the probability that $f(x) \neq f(y)$ when x is chosen uniformly at random and y is obtained from x by permuting the values of $\{x_i\}_{i \in S}$. This notion characterizes partially symmetric functions and satisfies several other useful properties. We provide the details in Section III.

The proof of the junta testing result in [29] relies on nice properties of the Fourier representation of the notion of influence. While symmetric influence also has a clean Fourier representation, it unfortunately does not have the properties needed to carry over the proof in [29] to the setting of partially symmetric functions. Instead, we must come up with a new proof technique.

Our proof of Theorem 2 uses a new connection to intersecting families. A family $\mathcal{F}$ of subsets of $[n]$ is t-*intersecting* if for every pair of sets $S, T \in \mathcal{F}$, their intersection size is at least $|S \cap T| \geq t$. This notion was introduced by Erdős, Ko, and Rado and a sequence of works led to the complete characterization of the maximum size of t-intersecting families that contain sets of fixed size [30], [31], [32], [33]. Dinur, Safra, and Friedgut recently extended those results to give bounds on the biased measure of intersecting families [34], [35].

Using results in intersecting families, we obtain a new and improved proof for the main lemma at the heart of the junta testing result [29]. The new proof is the first purely combinatorial analysis of a junta testing algorithm, as all previous proofs [9], [29] used Fourier-analytic arguments.[3] We describe the new proof and the connection to intersecting families in Section II. The same technique can also be extended to complete the proof of Theorem 2. We present this proof in Section IV.

The second and final component of the isomorphism test for partially symmetric functions is an efficient way to sample the core of such functions. An $(n-k)$-symmetric function f, which is symmetric over the complement of a set $J \subseteq [n]$ of size $|J| = k$, has a concise representation as a function $f_{\mathrm{core}} : \{0,1\}^k \times \{0,1,\ldots,n-k\} \to \{0,1\}$ which we call the *core* of f. The core is the restriction of f to the variables in J (in the natural order), with the additional Hamming weight of the variables outside of J. To determine if two partially symmetric functions are isomorphic, it suffices to determine whether their cores are isomorphic. We do so with the help of an efficient sample extractor.

[3]While Friedgut [35] used tools from Fourier analysis to bound the biased measure of intersecting families, he did so in order to obtain stability results that we do not use in this paper. The result that we use, stated below in Theorem 4, is easily obtained by extending Dinur and Safra's purely combinatorial argument [34].

Definition 3. A (1 query) δ-*sampler* for the $(n-k)$-symmetric function $f : \{0,1\}^n \to \{0,1\}$ is a randomized algorithm that queries f on a single input and returns a triplet $(x, w, z) \in \{0,1\}^k \times \{0,1,\ldots,n-k\} \times \{0,1\}$ where

- The distribution of (x, w) is δ-close, in total variation distance, to x being uniform over $\{0,1\}^k$ and w being binomial over $\{0,1,\ldots,n-k\}$ independently, and
- $z = f_{\mathrm{core}}(x, w)$ with probability at least $1-\delta$.

Our third main result is that for any $(n-k)$-symmetric function f, there is a query-efficient algorithm for constructing a δ-sampler for f.

Theorem 3. *Let $f : \{0,1\}^n \to \{0,1\}$ be $(n-k)$-symmetric with $k < n/10$. There is an algorithm that queries f on $O(\frac{k}{\eta\delta}\log\frac{k}{\eta\delta})$ inputs and with probability at least $1-\eta$ outputs a δ-sampler for f.*

This theorem is a generalization of a recent result of Chakraborty et al. [36], who gave a similar construction for sampling the core of juntas. Their result has many applications related to testing by implicit learning [37]. Our result may be of independent interest for similar such applications. We elaborate on this topic and present the proof of Theorem 3 in Section V.

D. Parallel and subsequent work

Chakraborty et al. [28] independently and simultaneously obtained a different proof that testing isomorphism to partially symmetric functions can be done with a constant number of queries. Their proof is significantly different than ours. The key to their argument is a clever reduction from the problem of testing partial symmetry to testing juntas. Thus, instead of having to generalize the junta testing algorithm (as we do in the current paper), they are able to use it as a black box to obtain an efficient partial symmetry tester. Our approach has a couple advantages. Notably, we obtain a nearly optimal bound of $O(k \log k)$ queries for testing k-symmetry, whereas the result in [28] gives a weaker $O(k^4 \log k)$ bound for the same task.

Another advantage of our approach is that the notion of symmetric influence, introduced in Section III and a key component of our analysis, appears to be a valuable tool for the study of partially symmetric functions in other contexts. Indeed, since the completion of the current work, Alon and Weinstein [38] have used symmetric influence in the analysis of a new algorithm for the local correction of partially symmetric functions.

II. Intersecting Families and Testing Juntas

We begin by revisiting the problem of junta testing. In this section, we give a new proof of the correctness of the k-junta tester first introduced in [29]. At a high level, the junta tester is quite simple: it partitions the set of indices into a large enough number of parts, then tries to identify all the parts that contain a relevant variable. If at most k such parts are found, the test accepts; otherwise it rejects. The algorithm is described in JUNTA-TEST.[4] In the algorithm and the discussion that follows, given a set $J \subseteq [n]$ and inputs $x, y \in \{0,1\}^n$, we write $x_J y_{\overline{J}}$ to represent the vector $z \in \{0,1\}^n$ that satisfies $z_i = x_i$ for each $i \in J$ and $z_i = y_i$ for each $i \in [n] \setminus J$.

Algorithm JUNTA-TEST(f, k, ϵ)

1: Create a random partition $\mathcal{I}$ of the set $[n]$ into $r = \Theta(k^2)$ parts, and initialize $J = \emptyset$.
2: **for** each $i = 1$ to $\Theta(k/\epsilon)$ **do**
3:     Sample $x, y \in \{0,1\}^n$ uniformly at random.
4:     **if** $f(x) \neq f(x_J y_{\overline{J}})$ **then**
5:         Use binary search to find a set $I \in \mathcal{I}$ that contains a relevant variable.
6:         Set $J := J \cup I$.
7:         **if** J is the union of $> k$ parts **then** reject.
8: Accept.

It is clear that the JUNTA-TEST always accepts k-juntas. The non-trivial part of the analysis involves showing that functions that are far from k-juntas are rejected by the tester with sufficiently high probability. To do so, we must argue that the inequality in Step 4 is satisfied with non-negligible probability whenever f is far from k-juntas and J is the union of at most k parts. This is accomplished by considering the influence of variables in a function.

The *influence* of the set $J \subseteq [n]$ in $f : \{0,1\}^n \to \{0,1\}$ is $\mathrm{Inf}_f(J) := \Pr_{x,y}[f(x) \neq f(x_{\overline{J}} y_J)]$. By definition, the probability that the inequality in Step 4 is satisfied is exactly $\mathrm{Inf}_f(\overline{J})$. To complete the analysis of correctness of the algorithm, we want to show that when f is ϵ-far from k-juntas, then with high probability over the choice of the random partition $\mathcal{I}$, for every set J obtained by taking the union of at most k parts in $\mathcal{I}$, $\mathrm{Inf}_f(\overline{J}) \geq \frac{\epsilon}{4}$. We do so by exploiting only a couple basic facts about the notion of influence.

Lemma 1 (Fischer et al. [9])**.** *For every $f : \{0,1\}^n \to \{0,1\}$ and every $J, K \subseteq [n]$, $\mathrm{Inf}_f(J) \leq \mathrm{Inf}_f(J \cup K) \leq \mathrm{Inf}_f(J) + \mathrm{Inf}_f(K)$. Also, if f is ϵ-far from k-juntas and $|J| \leq k$, then $\mathrm{Inf}_f(\overline{J}) \geq \epsilon$.*

We also use the fact that when f is far from k-juntas, the family of sets $J \subseteq [n]$ whose complements have small influence in f is an intersecting family. For a fixed $t \geq 1$, a family $\mathcal{F}$ of subsets of $[n]$ is called *t-intersecting* if any two sets J and K in $\mathcal{F}$ have intersection size $|J \cap K| \geq t$. Much of the work in this area focused on bounding the size of t-intersecting families that contain only sets of a fixed size. Dinur and Safra [34] considered general families and asked what the maximum *p-biased measure* of such families can be. For $0 < p < 1$, this measure is defined as $\mu_p(\mathcal{F}) := \Pr_J[J \in \mathcal{F}]$ where the probability over J is obtained by including each coordinate $i \in [n]$ in J independently with probability p. They showed that 2-intersecting families have small p-biased measure [34] and Friedgut showed how the same result also extends to t-intersecting families for $t > 2$ [35].

Theorem 4 (Dinur and Safra [34]; Friedgut [35])**.** *Let $\mathcal{F}$ be a t-intersecting family of subsets of $[n]$ for some $t \geq 1$. For any $p < \frac{1}{t+1}$, the p-biased measure of $\mathcal{F}$ is bounded by $\mu_p(\mathcal{F}) \leq p^t$.*

We are now ready to complete the analysis of JUNTA-TEST.

Lemma 2. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function that is ϵ-far from k-juntas and $\mathcal{I}$ be a random partition of $[n]$ into $r = 20k^2$ parts. Then with probability at least $5/6$, $\mathrm{Inf}_f(\overline{J}) \geq \epsilon/4$ for any union J of k parts from $\mathcal{I}$.*

Proof: For $0 \leq t \leq \frac{1}{2}$, let $\mathcal{F}_t = \{J \subseteq [n] : \mathrm{Inf}_f(\overline{J}) < t\epsilon\}$ be the family of all sets whose complements have influence at most $t\epsilon$. For any two sets $J, K \in \mathcal{F}_{1/2}$, the sub-additivity of influence implies that

$$\mathrm{Inf}_f(\overline{J \cap K}) = \mathrm{Inf}_f(\overline{J} \cup \overline{K}) \leq \mathrm{Inf}_f(\overline{J}) + \mathrm{Inf}_f(\overline{K}) < \epsilon \ .$$

But f is ϵ-far from k-juntas, so every set $S \subseteq [n]$ of size $|S| \leq k$ satisfies $\mathrm{Inf}_f(\overline{S}) \geq \epsilon$. Therefore, $|J \cap K| > k$ and, since this argument applies to every pair of sets in the family, $\mathcal{F}_{1/2}$ is a $(k+1)$-intersecting family.

Let us now consider two separate cases: when $\mathcal{F}_{1/2}$ contains a set of size less than $2k$; and when it does not. In the first case, let $J \in \mathcal{F}_{1/2}$ be one of the sets of size $|J| < 2k$. With high probability, the set J is completely separated by the partition $\mathcal{I}$, i.e., each element of J occupies a distinct part of $\mathcal{I}$. When this event occurs, then for every other set $K \in \mathcal{F}_{1/2}$, the fact that $|J \cap K| \geq k+1$ implies that K is not covered by any union of k parts in $\mathcal{I}$. Therefore, $\mathrm{Inf}_f(\overline{J}) \geq \frac{\epsilon}{2} > \frac{\epsilon}{4}$ for any union J of k parts from $\mathcal{I}$, as we wanted to show.

Consider now the case where $\mathcal{F}_{1/2}$ contains only sets of size at least $2k$. Then we claim that $\mathcal{F}_{1/4}$ is a $2k$-intersecting family: otherwise, we could find sets $J, K \in \mathcal{F}_{1/4}$ such that $|J \cap K| < 2k$ and $\mathrm{Inf}_f(\overline{J \cap K}) \leq \mathrm{Inf}_f(\overline{J}) + \mathrm{Inf}_f(\overline{K}) < \frac{\epsilon}{2}$, contradicting our assumption.

Let $J \subseteq [n]$ be the union of k parts in $\mathcal{I}$. Since $\mathcal{I}$ is a random partition, J is a random subset obtained by including each element of $[n]$ in J independently with probability $p = \frac{k}{r} < \frac{1}{2k+1}$. By Theorem 4, $\Pr_{\mathcal{I}}[\mathrm{Inf}_f(\overline{J}) < \frac{\epsilon}{4}] = \Pr[J \in \mathcal{F}_{1/4}] = \mu_{k/r}(\mathcal{F}_{1/4}) \leq (k/r)^{2k}$. By the union bound, the probability that there exists a set $J \subseteq [n]$ that is the union of k parts in $\mathcal{I}$ for which $\mathrm{Inf}_f(\overline{J}) < \frac{\epsilon}{4}$ is bounded above by $\binom{r}{k}\left(\frac{k}{r}\right)^{2k} \leq \left(\frac{er}{k}\right)^k \left(\frac{k}{r}\right)^{2k} \leq \left(\frac{ek}{r}\right)^k = \left(\frac{e}{20k}\right)^k < \frac{1}{6}$. ■

[4]See also [29] for more details on this algorithm.

III. SYMMETRIC INFLUENCE

The main focus of this paper is partially symmetric functions, that is, functions invariant under any reordering of the variables of some set $J \subseteq [n]$. Let $\mathcal{S}_J$ denote the set of permutations of $[n]$ which only move elements from the set J. A function $f : \{0,1\}^n \to \{0,1\}$ is *J-symmetric* if $f(x) = f(\pi x)$ for every input x and a permutation $\pi \in \mathcal{S}_J$, where πx is the vector whose $\pi(i)$-th coordinate is x_i.

To analyze partially symmetric functions, we introduce a new measure named *symmetric influence*. The symmetric influence of a set of coordinates measures the sensitivity of a function to random permutations of the labels of those coordinates.

Definition 4. The *symmetric influence* of a set $J \subseteq [n]$ of variables in a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is defined as $\mathrm{SymInf}_f(J) = \Pr_{x \in \{0,1\}^n, \pi \in \mathcal{S}_J}[f(x) \neq f(\pi x)]$.

It is not hard to see that a function f is t-symmetric iff there exists a set J of size t such that $\mathrm{SymInf}_f(J) = 0$. A much stronger connection, however, exists between these properties as we will shortly describe.

Before showing some nice properties of symmetric influence, we mention that it also has a simple representation using Fourier coefficients of the function. Although we do not use the representation in this paper, we feel it might be of independent interest. We describe this connection in more details in the full version of the paper.

Lemma 3. *Fix $f : \{0,1\}^n \to \{0,1\}$ and $J \subseteq [n]$. Let f_J be the J-symmetric function closest to f. The symmetric influence of J satisfies*

$$\mathrm{dist}(f, f_J) \leq \mathrm{SymInf}_f(J) \leq 2 \cdot \mathrm{dist}(f, f_J).$$

Proof: For every weight $0 \leq w \leq n$ and $z \in \{0,1\}^{|\overline{J}|}$, define the layer $L^w_{\overline{J} \leftarrow z} := \{x \in \{0,1\}^n \mid |x| = w \wedge x_{\overline{J}} = z\}$ to be the vectors of Hamming weight w which identify with z over the set $\overline{J}$ (where $|L^w_{\overline{J} \leftarrow z}| = \binom{|J|}{w-|z|}$ if $|z| \leq w \leq |J| + |z|$ or 0 otherwise). Let $p^w_z \in [0, \frac{1}{2}]$ be the fraction of the vectors in $L^w_{\overline{J} \leftarrow z}$ one has to modify in order to make the restriction of f over $L^w_{\overline{J} \leftarrow z}$ constant.

With this notation, we can restate the definition of the symmetric influence of J as follows.

$$\begin{aligned}\mathrm{SymInf}_f(J) =& \sum_z \sum_w \Pr_{x \in \{0,1\}^n}[x \in L^w_{\overline{J} \leftarrow z}] \cdot \\ & \Pr_{x \in \{0,1\}^n, \pi \in \mathcal{S}_J}[f(x) \neq f(\pi x) \mid x \in L^w_{\overline{J} \leftarrow z}] \\ =& \frac{1}{2^n} \sum_z \sum_w |L^w_{\overline{J} \leftarrow z}| \cdot 2p^w_z(1 - p^w_z)\ .\end{aligned}$$

The last identity holds because in each layer, the probability that x and πx result in two different outcomes is the probability that x is chosen out of the smaller part and πx from the complement, or vice versa.

The function f_J can be obtained by modifying f at p^w_z fraction of the inputs in each layer $L^w_{\overline{J} \leftarrow z}$, since each layer can be addressed separately and we want to modify as few inputs as possible. By this observation, we have that $\mathrm{dist}(f, f_J) = \frac{1}{2^n} \sum_z \sum_w |L^w_{\overline{J} \leftarrow z}| \cdot p^w_z$. Since $1 - p^w_z \in [\frac{1}{2}, 1]$, we have that $p^w_z \leq 2p^w_z(1 - p^w_z) \leq 2p^w_z$ and therefore $\mathrm{dist}(f, f_J) \leq \mathrm{SymInf}_f(J) \leq 2 \cdot \mathrm{dist}(f, f_J)$. ■

Corollary 1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function that is ϵ-far from being t-symmetric. Then every set $J \subseteq [n]$ of size $|J| \geq t$ has symmetric influence $\mathrm{SymInf}_f(J) \geq \epsilon$.*

Proof: Fix $J \subseteq [n]$ of size $|J| \geq t$ and let g be a J-symmetric function closest to f. Since g is symmetric on any subset of J, it is in particular t-symmetric and therefore $\mathrm{dist}(f, g) \geq \epsilon$ as f is ϵ-far from being t-symmetric. Thus, by Lemma 3, $\mathrm{SymInf}_f(J) \geq \mathrm{dist}(f, g) \geq \epsilon$ holds. ■

Corollary 1 demonstrates the strong connection between symmetric influence and the distance from being partially symmetric, similar to the second part of Lemma 1 for influence and juntas. The additional properties of influence used in Section II are monotonicity and sub-additivity (Lemma 1). The following lemmas show that the same properties approximately hold for symmetric influence. The proofs of these lemmas are in the full version of the paper.

Lemma 4 (Monotonicity)**.** *For any function $f : \{0,1\}^n \to \{0,1\}$ and any sets $J \subseteq K \subseteq [n]$,*

$$\mathrm{SymInf}_f(J) \leq \mathrm{SymInf}_f(K).$$

Lemma 5 (Weak sub-additivity)**.** *There is a universal constant c such that for any constant $0 < \gamma < 1$, any function $f : \{0,1\}^n \to \{0,1\}$, and any sets $J, K \subseteq [n]$ of size at least $(1 - \gamma)n$,*

$$\mathrm{SymInf}_f(J \cup K) \leq \mathrm{SymInf}_f(J) + \mathrm{SymInf}_f(K) + c\sqrt{\gamma}.$$

Note that symmetric influence does *not* satisfy the (strong) sub-additivity property. For example, consider the function $f(x) = f_1(x_J) \oplus f_2(x_K)$ where J and K partition $[n]$ and where f_1, f_2 are symmetric functions. While $\mathrm{SymInf}_f(J) = \mathrm{SymInf}_f(K) = 0$, the function f may be far from symmetric, in which case $\mathrm{SymInf}_f([n]) = \mathrm{SymInf}_f(J \cup K) > 0$.

IV. TESTING PARTIAL SYMMETRY

Let us now return to the problem of testing partial symmetry. The goal of this section is to introduce an efficient tester for this property by combining the ideas from Sections II and III.

We first introduce the testing algorithm PARTIALLY-SYMMETRIC-TEST. This algorithm is conceptually similar to the junta tester in Section II. Again, the main idea is to partition the variables into $O(k^2)$ parts and identify the parts that contain "asymmetric" variables. More precisely, given a function $f : \{0,1\}^n \to \{0,1\}$, let $J \subseteq [n]$ be the minimum set of variables such that f is $\overline{J}$-symmetric. We call the variables in J *asymmetric* and the variables in $[n]\setminus J$ are called *symmetric*. A function is $(n-k)$-symmetric iff it contains at most k asymmetric variables. The algorithm exploits this characterization by trying to identify $k+1$ parts that contain asymmetric variables.

Notice that unlike the tester for juntas, the Hamming weight of our queries plays an important role. Therefore, we dedicate one of the parts in our random partition to be a *workspace*, which we hope will not contain any asymmetric variables. We use the workspace to maintain the Hamming weight constant while modifying our query gradually to identify an additional part with an asymmetric variable.

Algorithm PARTIALLY-SYMMETRIC-TEST(f, k, ϵ)

Create a random partition $\mathcal{I}$ of $[n]$ into $r = \Theta(k^2/\epsilon^2)$ parts, and initialize $J := \emptyset$.
Pick a random workspace $W \in \mathcal{I}$, and **if** $|W| < \frac{n}{2r}$ **then** fail.
for each $i = 1$ to $\Theta(k/\epsilon)$ **do**
  Let $I :=$ FIND-ASYMMETRIC-SET$(f, \mathcal{I}, J, W)$.
  if $I \neq \emptyset$ **then**
    Set $J := J \cup I$.
    if J is the union of $> k$ parts **then** reject.
Accept.

The idea behind the FIND-ASYMMETRIC-SET algorithm is as follows. Suppose that we have two inputs $x, y \in \{0,1\}^n$ with $x_J = y_J, |x| = |y|$ such that $f(x) \neq f(y)$. Given such inputs, we know there exists some asymmetric variable outside of J. In order to efficiently find a set from a partition $\mathcal{I}$ which contains such a variable, we use binary search over the sets. First, we construct a refinement $\mathcal{J}$ of $\mathcal{I}$. Every set of $\mathcal{I}\setminus\{W\}$ is partitioned further into parts so that each part has size at most $\lceil |W|/4 \rceil$. Let $t = |\mathcal{J} \setminus \{W\}|$ be the number of parts in $\mathcal{J}$ excluding the workspace. Notice that the number of parts is at most $t \leq r + 4n/|W| = O(r)$. Then, we construct a series of inputs $x^0 = x, x^1, \ldots, x^t = y$ by each step permuting only elements from some set $I \in \mathcal{J} \setminus \{W\}$ and the workspace W (that is, applying a permutation from $\mathcal{S}_{I \cup W}$). In each such step, we guarantee that $x^i_I = y_I$ for one more set $I \in \mathcal{J} \setminus \{W\}$, and therefore after (at most) t steps we would reach y (notice that we can choose the last step such that $x^t_W = y_W$ as the Hamming weight of all the inputs in the sequence is identical).

Using this construction, we can now describe the algorithm FIND-ASYMMETRIC-SET as follows.

The following analysis of the FIND-ASYMMETRIC-SET algorithm shows that it satisfies the properties we need for testing partial symmetry.

Lemma 6. *Let f be a function, let $\mathcal{I}$ be a partition of $[n]$ into r parts, let $W \in \mathcal{I}, |W| \geq \frac{n}{2r}$ be a workspace, and let J be a union of parts from $\mathcal{I} \setminus \{W\}$. Then* FIND-ASYMMETRIC-SET$(f, \mathcal{I}, J, W)$ *performs $O(\log r)$ queries and*

Algorithm FIND-ASYMMETRIC-SET$(f, \mathcal{I}, J, W)$

Generate $x \in \{0,1\}^n$ and $\pi \in \mathcal{S}_{\overline{J}}$ uniformly at random.
if $f(x) \neq f(\pi x)$ **then**
  Define $x^0, \ldots, x^t$.
  Perform binary search on $x = x^0, \ldots, x^t = y$, and find i such that $f(x^{i-1}) \neq f(x^i)$.
  return the only part $I \in \mathcal{I} \setminus \{W\}$ such that $x^{i-1}_I \neq x^i_I$.
return $\emptyset$.

1) *With probability* $\mathrm{SymInf}_f(\overline{J})$*, it returns a set $I \in \mathcal{I} \setminus \{W\}$ disjoint to J; otherwise it returns $\emptyset$.*
2) *If W has no asymmetric variable and $I \in \mathcal{I}$ is returned, then I contains an asymmetric variable.*

Proof: Since we perform binary search over the sequence $x^0, \ldots, x^t$, the query complexity of the algorithm is indeed $O(\log t) = O(\log r)$. Also, it is easy to verify that we only output an empty set or a part in $\mathcal{I} \setminus \{W\}$ disjoint to J (since $x_J = y_J$).

Two random inputs x and $y := \pi x$, for $\pi \in \mathcal{S}_J$, satisfy $f(x) \neq f(y)$ with probability $\mathrm{SymInf}_f(\overline{J})$. Thus, it suffices to show that we can always define a sequence of $x^0, \ldots, x^t$, given that $|W| \geq \frac{n}{2r}$. In order to see that this is always feasible, we consider the sequence after already defining $x^0, \ldots, x^i$, and we show that we can define x^{i+1}.

Let $\mathcal{J}^+ = \{I \in \mathcal{J} \mid |x^i_I| > |y_I|\}$ and $\mathcal{J}^- = \{I \in \mathcal{J} \mid |x^i_I| < |y_I|\}$ denote the sets which require increasing or decreasing the Hamming weight of x_W respectively, when applying a permutation from $\mathcal{S}_{I \cup W}$ to ensure $x^{i+1}_I = y_I$. Notice that we ignore sets I for which $|x^i_I| = |y_I|$, as they do not impact the Hamming weight of x^i_W. If $|\mathcal{J}^+| > 0$ and $|\mathcal{J}^-| > 0$, then since $\max(|x^i_W|, |W| - |x^i_W|) \geq \lceil |W|/2 \rceil$ and the size of every set $I \in \mathcal{J} \setminus \{W\}$ is at most $\lceil |W|/4 \rceil$, there must exists a set we can use to define x^{i+1}. On the other hand, if $|\mathcal{J}^+| = 0$ for example, then we can define x^{i+1} using any set from $\mathcal{J}^-$ as $|x^i_W| - |y_W| = -\sum_{I \in \mathcal{J}\setminus\{W\}} |x^i_I| - |y_I|$ (recall that $|x| = |x^i| = |y|$).

It remains to show that when W contains no asymmetric variables and we output a part $I \in \mathcal{I} \setminus \{W\}$, I contains an asymmetric variable. Suppose that the output I is the part which was modified between x^{i-1} and x^i. Then, since $f(x^{i-1}) \neq f(x^i), |x^{i-1}| = |x^i|$, and x^{i-1} and x^i differ only on $I \cup W$, an asymmetric variable exists in $I \cup W$ and we know it is not in W. ■

Another important challenge in the analysis of PARTIALLY-SYMMETRIC-TEST is the use of symmetric influence (rather than influence). Similar to Lemma 2 for influence, we prove that if a function is far from being $(n-k)$-symmetric, then it is also far from being symmetric on any union of all but k parts of a random partition (assuming it has enough parts). The formal statement is given in Lemma 7.

Lemma 7. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function that is ϵ-far from $(n-k)$-symmetric and $\mathcal{I}$ be a random partition of $[n]$ into $r = c{\cdot}k^2/\epsilon^2$ parts, for some large enough constant c. Then with probability at least $8/9$, $\mathrm{SymInf}_f(\overline{J}) \geq \frac{\epsilon}{9}$ holds for any union J of k parts.*

The proof of this lemma is very similar to that of Lemma 2. The main difference between the two proofs is due to the weak-subadditivity of symmetric influence (compared to the subadditivity of influence). In light of this difference, our definition of families of sets whose complement has small symmetric influence includes only sets which are not too big. We use the observation that adding sets which contain elements of a family does not change its existing intersection. In addition, due to the additive factor of the sub-additivity we prove a slightly weaker result where the symmetric influence is at least $\epsilon/9$ and not $\epsilon/4$. The complete proof of Lemma 7 appears in the full version of this article.

We now complete the proof that partial symmetry is efficiently testable.

Proof of Theorem 2: Note that $|W| \geq \frac{n}{2r}$ indeed holds with probability at least $8/9$ from the Chernoff bound. By Lemma 6, FIND-ASYMMETRIC-SET performs $O(\log \frac{k}{\epsilon})$ queries according to our choice of r, and therefore the query complexity of PARTIALLY-SYMMETRIC-TEST is $O(\frac{k}{\epsilon} \log \frac{k}{\epsilon})$.

Suppose f is an $(n-k)$-symmetric function. The probability that W contains an asymmetric variable is at most $k/r \leq 2/9$. Conditioned on this event not occurring, every set returned by FIND-ASYMMETRIC-SET contains an asymmetric variable. Since there are at most k such variables, J would be the union of at most k sets and we would accept.

Suppose f is a function that is ϵ-far from being $(n-k)$-symmetric. From Lemma 7, with probability at least $8/9$, $\mathrm{SymInf}_f(\overline{J}) \geq \epsilon/9$ holds while J consists of at most k parts. Conditioned on that, by executing FIND-ASYMMETRIC-SET $O(k/\epsilon)$ times we obtain more than k parts with probability at least $8/9$, according to Lemma 6. Thus, we reject with probability at least $2/3$. ■

V. ISOMORPHISM TESTING OF PARTIALLY SYMMETRIC FUNCTIONS

In this section we prove that isomorphism testing of partially symmetric functions can be done with a constant number of queries. The algorithm we describe consists of two main components, and follow a similar approach to the one used in [10] to show that juntas are isomorphism testable. The first component, which we already described in Section IV, is an efficient tester for the property of being partially symmetric. Once we know the input function is indeed close to being partially symmetric, we can verify it is isomorphic (or at least very close to isomorphic) to the target function. The second component of the algorithm is therefore an efficient sampler from the *core* of a function which is (close to) partially symmetric. Comparing the cores of two partially symmetric functions suffices to identify if two such functions are isomorphic or far from it.

Ideally, when sampling the core of a partially symmetric function f, we would like to sample it according to the marginal distribution of sampling f at a uniform input $x \in \{0,1\}^n$. We denote this marginal distribution over $\{0,1\}^k \times \{0,1,\ldots,n-k\}$ by $\mathcal{D}^*_{k,n}$, which is in fact uniform over $\{0,1\}^k$ and binomial over $\{0,1,\ldots,n-k\}$, independently.

In our scenario, sampling the core of a function according to this distribution is not possible since we do not know the exact location of all the k asymmetric variables. Instead, we use the knowledge discovered by the partial symmetry tester, i.e., sets with asymmetric variables. Given these sets, we are able to define a sampling distribution over $\{0,1\}^n$ such that we know the input of the core for each query, and whose marginal distribution over the core is close enough to $\mathcal{D}^*_{k,n}$.

Definition 5. Let $\mathcal{I}$ be some partition of $[n]$ into an odd number of parts and let $W \in \mathcal{I}$ be the workspace. Define the distribution $\mathcal{D}^W_{\mathcal{I}}$ over $\{0,1\}^n$ to be as follows. Pick a random Hamming weight w according to the binomial distribution over $\{0,\ldots,n\}$ and output, if it exists, a random $x \in \{0,1\}^n$ of Hamming weight $|x| = w$ such that for every part $I \in \mathcal{I} \setminus \{W\}$, either $x_I \equiv 0$ or $x_I \equiv 1$. When no such x exists, return the all zeros vector.

The sampling distribution which we just defined, together with the random choice of the partition and workspace, satisfies two important properties: it is close to uniform over the inputs of the function, and its marginal distribution over the core of a partially symmetric function close to $\mathcal{D}^*_{k,n}$. These properties are formally written here as Proposition 1, whose proof is rather technical and deferred to the full version.

Proposition 1. *Let $J = \{j_1,\ldots,j_k\} \subseteq [n]$ be a set of size k, and $r = \Omega(k^2)$ be odd. If $x \sim \mathcal{D}^W_{\mathcal{I}}$ for a random partition $\mathcal{I}$ of $[n]$ into r parts and a random workspace $W \in \mathcal{I}$, then*

- *x is $o(1/n)$-close to being uniform over $\{0,1\}^n$, and*
- *$(x_J, |x_{\overline{J}}|)$ is c/k-close to being distributed according to $\mathcal{D}^*_{k,n}$, for our choice of $0 < c < 1$.*

We are now ready to describe the algorithm for isomorphism testing of $(n-k)$-symmetric functions. Given an $(n-k)$-symmetric function f, the algorithm tests whether the input function g is isomorphic to f or ϵ-far from being so.

The analysis of the algorithm is based on the fact that functions which pass the PARTIALLY-SYMMETRIC-TEST satisfy some conditions, and particularly are close to being partially symmetric, as described the following lemma.

Lemma 8. *Let g be a function that is ϵ-close to being $(n-k)$-symmetric and that passed the* PARTIALLY-SYMMETRIC-

Algorithm PARTIALLY-SYMMETRIC-ISO-TEST(f, k, g, ϵ)

1: Perform PARTIALLY-SYMMETRIC-TEST$(g, k, \epsilon/1000)$ and reject if failed.
2: Let $\mathcal{I}$ and $W \in \mathcal{I}$ be the partition and workspace used by the algorithm.
3: Let J be the union of the k parts identified by the algorithm (adding arbitrary parts if needed).
4: **for** each $i = 1$ to $\Theta(k \log k/\epsilon^2)$ **do**
5: Query $g(x)$ at a random $x \sim \mathcal{D}_{\mathcal{I}}^{W}$.
6: Accept iff $(1-\epsilon/2)$-fraction of the queries are consistent with some isomorphism f_π of f where π maps the asymmetric variables of f into all k parts of J.

TEST(g, k, ϵ). *In addition, let $\mathcal{I}, W$ and J be the partition, workspace and identified parts used by the algorithm. With probability at least $9/10$, there exists a function h which satisfies the following properties.*

- *h is 4ϵ-close to g,*
- *h is $(n-k)$-symmetric, and*
- *the asymmetric variables of h are contained in J and separated by $\mathcal{I}$.*

Proof: Let g^* be the $(n-k)$-symmetric function closest to g (which can be f itself, up to some isomorphism) and let R be the set of (at most) k asymmetric variables of g^*. By Lemma 3 and our assumption on g,

$$\mathrm{SymInf}_g(\overline{R}) \leq 2 \cdot \mathrm{dist}(g, g^*) \leq 2\epsilon \ .$$

Notice however that R is not necessarily contained in J and therefore g^* may not be a good enough candidate for h. Let $U = R \cap J$ be the intersection of the asymmetric variables of g^* and the sets identified by the algorithm. In order to show that g is also close to being $\overline{U}$-symmetric, we bound $\mathrm{SymInf}_g(\overline{U})$ using Lemma 5 with the sets $\overline{R}$ and $\overline{J}$. Notice that since $|R| \leq k$ and $|J| \leq 2kn/r \leq \epsilon^2 n/c'$ for our choice of c', we can bound the error term (in the notation of Lemma 5) by $c\sqrt{\gamma} \leq c\sqrt{\epsilon^2/c'} \leq \epsilon$. We therefore have

$$\mathrm{SymInf}_g(\overline{U}) \leq \mathrm{SymInf}_g(\overline{R}) + \mathrm{SymInf}_g(\overline{J}) + \epsilon \leq 4\epsilon$$

where we know $\mathrm{SymInf}_g(\overline{J}) \leq \epsilon$ with probability at least $19/20$ as the algorithm did not reject.

By applying Lemma 3 again, we know there exists a $\overline{U}$-symmetric function h, whose distance to g is bounded by $\mathrm{dist}(g, h) \leq 4\epsilon$. Moreover, with probability at least $19/20$, all its asymmetric variables are completely separated by the partition $\mathcal{I}$ (and they were all identified as part of J). ∎

Given Lemma 8, we are now ready to analyze PARTIALLY-SYMMETRIC-ISO-TEST.

Proof of Theorem 1: Before analyzing the algorithm we just described, we consider the case where $k > n/10$. Since Theorem 2 does not hold for such k's, we apply the basic algorithm of $O(n \log n/\epsilon)$ random queries, which is applicable for testing isomorphism of any given function (since there are $n!$ possible isomorphisms, the random queries will rule out all of them with good probability, assuming we should reject). Since $k = \Omega(n)$, the complexity of this algorithm fits the statement of our theorem.

We first analyze the query complexity of the algorithm. The step of PARTIALLY-SYMMETRIC-TEST performs $O(\frac{k}{\epsilon} \log \frac{k}{\epsilon})$ queries, and therefore the majority of the queries are performed at the sampling stage, resulting in $O(k \log k/\epsilon^2)$ queries as required. In order to prove the correctness of the algorithm, we consider the following cases.

- g is ϵ-far from being isomorphic to f and $\epsilon/1000$-far from being $(n-k)$-symmetric.
- g is ϵ-far from being isomorphic to f but $\epsilon/1000$-close to being $(n-k)$-symmetric.
- g is isomorphic to f.

In the first case, with probability at least $9/10$, PARTIALLY-SYMMETRIC-TEST will reject and so will we, as required. We assume from this point on that PARTIALLY-SYMMETRIC-TEST did not reject, as it will only reject g which is isomorphic to f with probability at most $1/10$, and that we are not in the first case. Notice that these cases match the conditions of Lemma 8, and therefore from this point onward we assume there exists an h satisfying the lemma's properties (remembering we applied the algorithm with $\epsilon/1000$).

In order to bound the distance between h and g in our samples, we use Proposition 1, indicating

$$\Pr_{\mathcal{I}, W \in \mathcal{I}, x \sim \mathcal{D}_{\mathcal{I}}^{W}}[g(x) \neq h(x)] = \mathrm{dist}(g, h) + o(1/n) \ .$$

By Markov's inequality, with probability at least $9/10$, the partition $\mathcal{I}$ and the workspace W satisfy

$$\begin{aligned} \Pr_{x \sim \mathcal{D}_{\mathcal{I}}^{W}}[g(x) \neq h(x)] &\leq 10 \cdot \mathrm{dist}(g, h) + o(1/n) \\ &\leq 10 \cdot 4\epsilon/1000 + o(1/n) < \epsilon/20 \ . \end{aligned}$$

By Proposition 1, if we were to sample h according to $\mathcal{D}_{\mathcal{I}}^{W}$, it should be $\epsilon/20$-close to sampling its core (assuming the partition size is large enough). Combined with the distance between g and h in our samples, we expect our samples to be $\epsilon/20 + \epsilon/20 = \epsilon/10$ close to sampling h's core.

The last part of the proof consists of showing that the only way that there can be an almost consistent isomorphism of f is when g is isomorphic to f. Notice however that we care only for isomorphisms which map the asymmetric variables of f to the k sets of J. Therefore, the number of different isomorphisms we need to consider is $k!$.

Assume we are in the second case and g is ϵ-far from being isomorphic to f. Let f_π be some isomorphism of f. By our assumptions and Lemma 8,

$$\mathrm{dist}(f_\pi, h) \geq \mathrm{dist}(f_\pi, g) - \mathrm{dist}(g, h) \geq \epsilon - \epsilon/250 \ .$$

Each sample we perform is inconsistent with f_π with probability at least $\epsilon - \epsilon/250 - \epsilon/10 > 8\epsilon/9$. By the Chernoff bounds and the union bound, if we perform $q = O(k \log k/\epsilon^2)$ queries, we rule out all $k!$ possible isomorphisms with probability at least $9/10$ and reject the function as required.

On the other hand, if g is isomorphic to f, then we know there exists with probability at least $9/10$ some isomorphism f_π which maps the asymmetric variables of f into the sets of J, such that

$$\mathrm{dist}(f_\pi, h) \le \mathrm{dist}(f_\pi, g) + \mathrm{dist}(g, h) \le \epsilon/500 + \epsilon/250 \ .$$

Notice that we cannot assume that $\mathrm{dist}(f_\pi, g) = 0$ as the algorithm may not identify all the asymmetric sets, if some barely influence the output. Using arguments similar to the ones in the proof of Lemma 8, we can bound this distance by $\epsilon/500$.

For this isomorphism, with high probability much more than $(1-\epsilon/2)$-fraction of the queries are consistent and we therefore accept g, as we should. ■

As we outlined above, we in fact build an efficient sampler for the core of $(n-k)$-symmetric functions (or functions close to being so). Given the parts identified by PARTIALLY-SYMMETRIC-TEST, assuming it did not reject, we can sample the function's core by querying it at a single location, where the distribution over the core's inputs is close to $\mathcal{D}^*_{k,n}$. The algorithm and proof of Theorem 3 are deferred to the full version.

VI. DISCUSSION

Our result unifies the previous classes of functions that are efficiently isomorphism-testable. More importantly, we believe that the query complexity for testing f-isomorphism is determined by the partial symmetry of f. Specifically, let $k_\epsilon(f)$ be the smallest k such that the function f is ϵ-close to an $(n-k)$-symmetric function and $q_\epsilon(f)$ be the minimum query complexity for testing f-isomorphism with an error parameter ϵ. We raise the following conjecture, which is analogous to the result by Fischer on the isomorphism testability of graphs [39].

Conjecture 1. *There exist a constant c and functions $L_\epsilon(k), U_\epsilon(k)$ with $\lim_{k\to\infty} L_\epsilon(k) = \infty$ such that, for every function $f : \{0,1\}^n \to \{0,1\}$, we have $L_\epsilon(k_{c\epsilon}(f)) \le q_\epsilon(f) \le U_\epsilon(k_{\epsilon/c}(f))$.*

Using symmetric influence and the analysis tools developed in the current paper, we can show that the upper bound of the conjecture holds. The lower bound remains open, but it is consistent with all known hardness results on testing function isomorphism. In particular, by the result in [15], we know that testing f-isomorphism requires at least $\Omega(k)$ queries for *almost all* functions f that are ϵ-far from $(n-k)$-symmetric. A simple extension of the proof in [14] shows that for every $(n-k)$-symmetric function f that is ϵ-far from $(n-k+1)$-symmetric, testing f-isomorphism requires $\Omega(\log\log k)$ queries (assuming k/n is bounded away from 1).

Lastly, let us consider another natural definition of partial symmetry that encompasses both symmetric functions and juntas. The function $f : \{0,1\}^n \to \{0,1\}$ is *k-part symmetric* if there is a partition $\mathcal{I} = \{I_1, \ldots, I_k\}$ of $[n]$ such that f is invariant under any permutation π of $[n]$ where $\pi(I_i) = I_i$ for every $i = 1, \ldots, k$. One may be tempted to guess that k-part symmetric functions are efficiently isomorphism-testable. That is not the case, even when $k = 2$. To see this, consider the function $f(x) = x_1 \oplus x_2 \oplus \cdots \oplus x_{n/2}$. This function is 2-part symmetric, but testing isomorphism to f requires $\Omega(n)$ queries [12].

ACKNOWLEDGMENT

We thank Noga Alon, Per Austrin, Irit Dinur, Ehud Friedgut, and Ryan O'Donnell for useful discussions and valuable feedback.

A.W.'s research was supported in part by an ERC Advanced grant and by the Israeli Centers of Research Excellence (I-CORE) program.

Most of this research was completed while Y.Y. was at Kyoto University.

REFERENCES

[1] R. Rubinfeld and M. Sudan, "Robust characterizations of polynomials with applications to program testing," *SIAM Journal on Computing*, vol. 25, no. 2, pp. 252–271, 1996.

[2] O. Goldreich, S. Goldwasser, and D. Ron, "Property testing and its connection to learning and approximation," *Journal of the ACM*, vol. 45, no. 4, pp. 653–750, 1998.

[3] D. Ron, "Algorithmic and analysis techniques in property testing," *Foundations and Trends in Theoretical Computer Science*, vol. 5, pp. 73–205, 2010.

[4] R. Rubinfeld and A. Shapira, "Sublinear time algorithms," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 18, 2011, tR11-013.

[5] O. Goldreich, Ed., *Property Testing: Current Research and Surveys*, ser. LNCS. Springer, 2010, vol. 6390.

[6] N. Alon, E. Fischer, I. Newman, and A. Shapira, "A combinatorial characterization of the testable graph properties: It's all about regularity," *SIAM Journal on Computing*, vol. 39, pp. 143–167, 2009.

[7] A. Bhattacharyya, E. Grigorescu, and A. Shapira, "A unified framework for testing linear-invariant properties," in *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 478–487.

[8] T. Kaufman and M. Sudan, "Algebraic property testing: the role of invariance," in *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC)*, 2008, pp. 403–412.

[9] E. Fischer, G. Kindler, D. Ron, S. Safra, and A. Samorodnitsky, "Testing juntas," *Journal of Computer and System Sciences*, vol. 68, no. 4, pp. 753–787, 2004.

[10] S. Chakraborty, D. García-Soriano, and A. Matsliah, "Nearly tight bounds for testing function isomorphism," in *Proc. 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2011, pp. 1683–1702.

[11] O. Goldreich, "On testing computability by small width OBDDs," *Proc. 14th International Workshop on Randomization and Approximation Techniques in Computer Science*, pp. 574–587, 2010.

[12] E. Blais, J. Brody, and K. Matulef, "Property testing lower bounds via communication complexity," in *Proc. 26th Annual IEEE Conference on Computational Complexity (CCC)*, 2011, pp. 210–220.

[13] E. Blais and D. Kane, "Testing linear functions," 2011, manuscript.

[14] E. Blais and R. O'Donnell, "Lower bounds for testing function isomorphism," in *Proc. 25th Conference on Computational Complexity (CCC)*, 2010, pp. 235–246.

[15] N. Alon, E. Blais, S. Chakraborty, D. García-Soriano, and A. Matsliah, "Nearly tight bounds for testing function isomorphism," 2011, manuscript.

[16] N. Alon and E. Blais, "Testing boolean function isomorphism," *Proc. 14th International Workshop on Randomization and Approximation Techniques in Computer Science*, pp. 394–405, 2010.

[17] C. E. Shannon, "The synthesis of two-terminal switching circuits," *Bell System Technical Journal*, vol. 28, no. 1, pp. 59–98, 1949.

[18] R. F. Arnold and M. A. Harrison, "Algebraic properties of symmetric and partially symmetric boolean functions," *IEEE Transactions on Electronic Computers*, vol. EC-12, no. 3, pp. 244–251, 1963.

[19] L. Babai, R. Beals, and P. Takácsi-Nagy, "Symmetry and complexity," in *Proc. 24th Annual ACM Symposium on Theory of Computing*, 1992, pp. 438–449.

[20] P. Clote and E. Kranakis, "Boolean functions, invariance groups, and parallel complexity," *SIAM Journal on Computing*, vol. 20, pp. 553–590, 1991.

[21] S. Das and C. Sheng, "On detecting total or partial symmetry of switching functions," *IEEE Trans. on Computers*, vol. C-20, no. 3, pp. 352–355, 1971.

[22] C. Meinel and T. Theobald, *Algorithms and Data Structures in VLSI Design*. Springer, 1998.

[23] H. A. Nienhaus, "Efficient multiplexer realizations of symmetric functions," in *Southeastcon '81*, 1981, pp. 522–525.

[24] T. Pitassi and R. Santhanam, "Effectively polynomial simulations," in *Proc. 1st Symposium on Innovations in Computer Science (ICS)*, 2010, pp. 370–382.

[25] T. Sasao and P. Besslich, "On the complexity of mod-2l sum PLA's," *IEEE Transactions on Computers*, vol. 39, no. 2, pp. 262–266, 1990.

[26] D. Sieling, "Variable orderings and the size of OBDDs for random partially symmetric boolean functions," *Random Structures & Algorithms*, vol. 13, no. 1, pp. 49–70, 1998.

[27] S. S. Yau and C. K. Tang, "Universal logic modules and their applications," *Computers, IEEE Transactions on*, vol. C-19, no. 2, pp. 141–149, 1970.

[28] S. Chakraborty, E. Fischer, D. García-Soriano, and A. Matsliah, "Junto-symmetric functions, hypergraph isomorphism, and crunching," in *Proc. 27th Annual IEEE Conference on Computational Complexity (CCC)*, 2012.

[29] E. Blais, "Testing juntas nearly optimally," in *Proc. 41st Annual ACM Symposium on Theory of Computing (STOC)*, 2009, pp. 151–158.

[30] P. Erdős, C. Ko, and R. Rado, "Intersection theorems for systems of finite sets," *The Quarterly Journal of Mathematics*, vol. 12, no. 1, pp. 313–320, 1961.

[31] P. Frankl, "The Erdős-Ko-Rado theorem is true for $n = ckt$," in *Combinatorics (Proc. Fifth Hungarian Colloquium, Keszthely)*, vol. 1, 1976, pp. 365–375.

[32] R. M. Wilson, "The exact bound in the Erdős-Ko-Rado theorem," *Combinatorica*, vol. 4, no. 2–3, pp. 247–257, 1984.

[33] R. Ahlswede and L. H. Khachatrian, "The complete intersection theorem for systems of finite sets," *European Journal of Combinatorics*, vol. 18, pp. 125–136, 1997.

[34] I. Dinur and S. Safra, "On the hardness of approximating minimum vertex cover," *Annals of Mathematics*, vol. 162, no. 1, pp. 439–485, 2005.

[35] E. Friedgut, "On the measure of intersecting families, uniqueness and stability," *Combinatorica*, vol. 28, no. 5, pp. 503–528, 2008.

[36] S. Chakraborty, D. García-Soriano, and A. Matsliah, "Efficient sample extractors for juntas with applications," *Automata, Languages and Programming*, pp. 545–556, 2011.

[37] I. Diakonikolas, H. Lee, K. Matulef, K. Onak, R. Rubinfeld, R. Servedio, and A. Wan, "Testing for concise representations," in *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007, pp. 549–558.

[38] N. Alon and A. Weinstein, "Local correction with constant error rate," 2012, manuscript.

[39] E. Fischer, "The difficulty of testing for isomorphism against a graph that is given in advance," in *Proc. 36th Annual ACM Symposium on Theory of Computing (STOC)*. ACM, 2004, pp. 391–397.