If P $\neq$ NP then Some Strongly Noninvertible Functions are Invertible *

Lane A. Hemaspaandra [a,1]

[a] *Dept. of Computer Science, University of Rochester, Rochester, NY 14627, USA*

Kari Pasanen [b,2]

[b] *University of Jyväskylä and Nokia Networks, Jyväskylä, Finland*

Jörg Rothe [c,3]

[c] *Institut für Informatik, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany*

Abstract

Rabi, Rivest, and Sherman alter the standard notion of noninvertibility to a new notion they call strong noninvertibility, and show—via explicit cryptographic protocols for secret-key agreement ([RS93,RS97] attribute this protocol to Rivest and Sherman) and digital signatures [RS93,RS97]—that strongly noninvertible functions are very useful components in protocol design. Their definition of strong noninvertibility has a small twist ("respecting the argument given") that is needed to ensure cryptographic usefulness. In this paper, we show that this small twist has a consequence: Unless P = NP, some strongly noninvertible functions are invertible.

Key words: computational complexity, cryptography, one-way functions, associativity, strong noninvertibility

* Supported in part by grants NSF-INT-9815095/DAAD-315-PPP-gü-ab, NSF-CCF-0426761, DFG-RO-1202/9-1, and DFG-RO-1202/9-3, a Heisenberg Fellowship from the DFG, and the Alexander von Humboldt Foundation's TransCoop program.

Email address: kari.pasanen@starnet.fi (Kari Pasanen).

URLs: www.cs.rochester.edu/u/lane (Lane A. Hemaspaandra), ccc.cs.uni-duesseldorf.de/~rothe (Jörg Rothe).

[1] Work done in part while visiting Julius-Maximilians-Universität Würzburg and Heinrich-Heine-Universität Düsseldorf.

[2] Current affiliation/address: Starnet Systems, FIN-40100 Jyväskylä, Finland.

[3] Corresponding author. Work done in part while visiting the University of Rochester.

1 Introduction

Rabi, Rivest, and Sherman developed novel cryptographic protocols that require one-way functions with algebraic properties such as associativity (see [RS93,RS97] and the attributions and references therein, especially [She86] and [KRS88]). Motivated by these protocols, they initiated the study of two-argument (2-ary, for short) one-way functions in worst-case cryptography. To preclude certain types of attacks, their protocols require one-way functions that are not invertible in polynomial time even when the adversary is given not just the function's output but also one of the function's inputs. Calling this property of one-way functions "strong noninvertibility" (or "strongness," for short), they left as an open problem whether there is any evidence—e.g., any plausible complexity-theoretic hypothesis—ensuring the existence of one-way functions with all the properties the protocols require, namely ensuring the existence of total, commutative, associative one-way functions that are strongly noninvertible. This problem was recently solved by Hemaspaandra and Rothe [HR99] who show that if $\mathrm{P} \neq \mathrm{NP}$ then such one-way functions do exist.

Hemaspaandra and Rothe [HR99] write: "Rabi and Sherman [RS97] also introduce the notion of *strong* one-way functions—2-ary one-way functions that are hard to invert even if one of their arguments is given. Strongness implies one-way-ness." The latter sentence could be very generously read as meaning "strong, one-way functions" when it speaks of "strongness," especially since strongness alone, by definition, does not even require honesty, and without honesty the sentence quoted above would be provably, trivially, false. However, a more natural reading is that [HR99] is assuming that strongly noninvertible functions are always noninvertible. The main result of the present paper is that if $\mathrm{P} \neq \mathrm{NP}$ then this is untrue. So, even when one has proven a function to be strongly noninvertible, one must not merely assume that noninvertibility automatically holds (as it may not), but rather one must prove the function's noninvertibility. [4]

In the present paper, we study appropriately honest, polynomial-time computable 2-ary functions. We prove that if $\mathrm{P} \neq \mathrm{NP}$ then there exist strongly noninvertible such functions that are invertible (see Section 2 for precise definitions). To paint a full picture of what happens if $\mathrm{P} \neq \mathrm{NP}$, we show that if

[4] Since in the paper [HR99] only strong noninvertibility is explicitly proven, one might worry that the functions constructed in its proofs may be invertible. Fortunately, the constructions in the proofs in [HR99] do easily support and implicitly give noninvertibility as well; thus, all the claims of Hemaspaandra and Rothe [HR99] remain correct. Most crucially, on page 654 of [HR99], inverting the output $\langle x, x \rangle$ in polynomial time would give strings containing one witness for membership of x in the given set in $\mathrm{NP} - \mathrm{P}$ (if there are any such witnesses), which is impossible.

$P \neq NP$ then there exist appropriately honest, polynomial-time computable 2-ary functions that are noninvertible, yet not strongly noninvertible.

So, why is the result that if $P \neq NP$ then some strongly noninvertible functions are invertible possible? Let us informally explain. Let σ be a 2-ary function. We say σ is noninvertible if there is no polynomial-time inverter that, given an image element z of σ, outputs some preimage of z. We say σ is strongly noninvertible if even when, in addition to any image element z of σ, one argument of σ is given such that there exists another string with which this argument is mapped to z, computing one such other argument is not a polynomial-time task. Why does strongness alone not outright imply noninvertibility? One might be tempted to think that from some given polynomial-time inverter g witnessing the invertibility of σ one could construct polynomial-time inverters g_1 and g_2 such that g_i inverts σ in polynomial time even when the ith argument is given (see Definition 2.2 for the formal details). This approach does not work. In particular, it is not clear how to define g_1 when given an output z of σ and a first argument a that together with a corresponding second argument is mapped to z, yet a is not the first component of $g(z)$. In fact, our main theorem implies that, unless $P = NP$, *no* approach can in general accomplish the desired transformation from g to g_1.

But then, why don't we use a different notion of strongness that automatically implies noninvertibility? The answer is that the definitional subtlety that opens the door to the unexpected behavior is absolutely essential to the cryptographic protocols for which Rabi, Rivest, and Sherman created the notion in the first place. For example, suppose one were tempted to redefine "strongly noninvertible" with the following quite different notion: σ is "strongly noninvertible" if, given any image element z of σ and any one argument of σ such that there exists another string with which this argument is mapped to z, computing *any preimage of* z (as opposed to "any other argument respecting the argument given") is not a polynomial-time task. The problem with this redefinition is that it completely loses the core of why strongness precludes direct attacks against the protocols of Rabi, Rivest, and Sherman. It is difficult to explain why without giving here in full their protocols; also, this intuitive argument is not a formal claim, just as Rabi and Sherman's arguments in [RS93,RS97] are not formal proofs of security. (However, later in this section we will sketch a formal proof that strong noninvertibility is a *necessary* condition for security of the Rivest–Sherman secret-key agreement protocol, and that one of the two components of strong noninvertibility, namely noninvertibility with respect to the second argument, is a *necessary* condition for security of the Rabi–Sherman digital signatures protocol.) Rabi and Sherman's intuitive arguments crucially draw on the fact that the original definition of strong noninvertibility includes the "respecting the argument given" feature, and this dependence will be immediately clear to anyone who reads their protocols. In the paragraph after the following one, we will return

to this issue, and will briefly sketch their protocols and explain the nature of the dependence of those on strong noninvertibility and on noninvertibility in the second argument.

The alternate notion defined at the start of the preceding paragraph will be called "overstrongness," since it seems to be so restrictive as to fail to be useful in the cryptographic protocols of Rabi, Rivest, and Sherman. In order to indicate that overstrongness is not simply an alternate notion equivalent to strongness, we in this paper will prove that if P $\neq$ NP, then overstrongness is a properly more restrictive notion than strongness.

For the interested reader, we now explain in a bit more detail the protocols of Rabi, Rivest, and Sherman, and justify our above claims that strong noninvertibility is a necessary condition for security of the Rivest–Sherman secret-key agreement protocol and that one of the two components of strong noninvertibility, namely noninvertibility with respect to the second argument, is a necessary condition for security of the Rabi–Sherman digital signatures protocol. The text from here to the end of the section is not needed to understand the body of the paper and so readers not interested in these protocols and why strong noninvertibility is important to them may safely skip forward to the start of Section 2.

The secret-key agreement protocol, which is attributed to Rivest and Sherman by Rabi and Sherman [RS93,RS97], works as follows. Suppose that Alice and Bob, who are communicating via an insecure channel that is being eavesdropped on by Eve, have a total, strongly noninvertible, associative one-way function, σ. Alice starts by choosing two large random strings, x and y, keeps x secret, and computes $\sigma(x, y)$. She sends y and $\sigma(x, y)$ to Bob. Then Bob chooses some large random string, z, keeps z secret, computes $\sigma(y, z)$, and sends the value $\sigma(y, z)$ back to Alice. They now can each compute their joint secret key: Alice computes her key $k_A = \sigma(x, \sigma(y, z))$ and Bob computes his key $k_B = \sigma(\sigma(x, y), z)$. Since σ is associative, both keys are the same—$k_A = \sigma(x, \sigma(y, z)) = \sigma(\sigma(x, y), z) = k_B$—so the protocol indeed reaches agreement on a key.

It is not known whether this secret-key agreement protocol is secure (see the discussions in [RS93,RS97,HR99]). However, note that the assumption that σ is strongly noninvertible is crucial: If σ is not strongly noninvertible, then the protocol is obviously insecure. In particular, if σ is not strongly noninvertible, then the following attack by Eve reveals the secret key. (The reader may wish to defer reading the following attack until after reading the rigorous definition of strong noninvertibility, namely, Definition 2.2. The reason is that in this attack we refer to the notions of invertibility with respect to the first argument and invertibility with respect to the second argument, both of which are defined in that definition.) So, suppose that σ is not strongly noninvertible.

That means (see Definition 2.2) that it is invertible with respect to its first argument or invertible with respect to its second argument. Let us treat the case in which it is invertible with respect to the first argument (the other case's proof is completely analogous). Eve in this case, since she knows $\sigma(y, z)$ and y, can invert in the first argument. Note that we must *not* claim that this gives Eve z. That would hold if σ were one-to-one, but we have not assumed one-to-one-ness. What we can claim is something weaker, namely, that Eve by this inversion obtains in polynomial time a string z' such that $\sigma(y, z) = \sigma(y, z')$. But note that by associativity $\sigma(\sigma(x, y), z') = \sigma(x, \sigma(y, z'))$. So, from the previous two sentences we conclude that $\sigma(\sigma(x, y), z') = \sigma(x, \sigma(y, z')) = \sigma(x, \sigma(y, z))$. But the rightmost member of this equality chain, $\sigma(x, \sigma(y, z))$, *is* the secret key. And the leftmost member of this equality chain, $\sigma(\sigma(x, y), z')$, is something that Eve can easily compute, since she has both z' and $\sigma(x, y)$. So Eve indeed can obtain the secret key.

Rabi and Sherman [RS93,RS97] modified the Rivest–Sherman protocol for secret-key agreement—which we just summarized and discussed—to obtain a protocol for digital signatures. The nature of the Rabi–Sherman modification is such that if the underlying function is invertible with respect to the second argument then their digital signature scheme is insecure. To make this clear, let us present their scheme and discuss how it relates to issues of invertibility in the second argument.

Alice starts by choosing two large random strings, x and y, keeps x secret, and computes $\sigma(x, y)$. She sends y and $\sigma(x, y)$ to a trusted third party who certifies that these values originate from her and makes them public (available to everyone). Suppose Alice wishes to sign a message m. Then she computes its signature, $\mathrm{sig}_A(m) = \sigma(m, x)$, and sends both m and $\mathrm{sig}_A(m)$ to the recipient, call him Bob. In this protocol, Bob's attempt to verify Alice's signature is as follows: Suppose Bob receives a message, purportedly from Alice, with the message being h and the signature being s. Bob recovers from the public database the y and the $\sigma(x, y)$ associated with Alice, and then Bob checks whether $\sigma(s, y)$ equals $\sigma(h, \sigma(x, y))$, and if so he considers it a valid signature and if not he considers it not to be a valid signature. Note that since σ is associative we have that if s is Alice's actual signature for h, then Bob certainly will consider the message to be valid, since $\sigma(s, y) = \sigma(\sigma(h, x), y) = \sigma(h, \sigma(x, y))$.

It is not known whether this secret-key agreement protocol is secure (see the discussions in [RS93,RS97,HR99]). However, it certainly is critical that σ be assumed to be noninvertible with respect to the second argument: If σ is invertible with respect to the second argument, then the protocol is obviously insecure. Eve in this case, since she knows $\sigma(x, y)$ and y, can invert in the second argument, so she by this inversion obtains in polynomial time a string x' such that $\sigma(x', y) = \sigma(x, y)$. Note that to get this, she didn't even need to intercept a message: Purely from the public database and invertibility in the

second argument, she was able to obtain x'. However, x' allows Eve to forge signatures for arbitrary messages to purportedly be from Alice. In particular, if Eve wants to forge a signature for an arbitrary message m', she simply sends out the message m' and the signature $\sigma(m', x')$. This forged signature might well not be equal to the signature Alice would have used to send the same message, namely $\sigma(m', x)$, so perhaps Alice will not be fooled (even aside from the fact that Alice might recognize m' to be something she never sent, if it happens to be an m' she never sent). But the goal is to fool the verification algorithm, and *that* will be fooled if shown the message. It will be fooled because

$$\sigma(\sigma(m', x'), y) = \sigma(m', \sigma(x', y)) = \sigma(m', \sigma(x, y)),$$

where the first equality holds due to associativity, and the second holds due to x' being from Eve's inversion. Thus we have shown that invertibility with respect to the second argument is a sufficient condition for the insecurity of this digital-signature scheme.

2 Definitions

Fix the binary alphabet $\Sigma = \{0, 1\}$. Let ϵ denote the empty string. Let $\langle \cdot, \cdot \rangle : \Sigma^* \times \Sigma^* \to \Sigma^*$ be some standard pairing function, that is, some total, polynomial-time computable bijection that has polynomial-time computable inverses and is nondecreasing in each argument when the other argument is fixed. Let FP denote the set of all polynomial-time computable total functions. The standard definition of one-way-ness used here is essentially due to Grollmann and Selman except that they require one-way functions to be one-to-one ([GS88], see also [Ko85,Ber77] and the surveys [Sel92,BHHR99], and see [All86,AR88] regarding the case of polynomial-to-one one-way functions). As in the papers [RS97,HR99,Hom04], Grollmann and Selman's notion of one-way-ness is tailored below to the case of 2-ary functions. Any general notions not explicitly defined can be found in standard complexity texts [BC93,Pap94, BDG95].

Definition 2.1 [GS88,RS97,HR99] *Let $\rho : \Sigma^* \times \Sigma^* \to \Sigma^*$ be any (possibly nontotal, possibly many-to-one) 2-ary function.*

(1) We say ρ is honest *if and only if there exists a polynomial q such that:*

$$(\forall z \in \text{image}(\rho))\,(\exists (a, b) \in \text{domain}(\rho))\,[|a| + |b| \leq q(|z|) \wedge \rho(a, b) = z].$$

(2) We say ρ is (polynomial-time) noninvertible *if and only if the following does not hold:*

$$(\exists g \in \text{FP})\,(\forall z \in \text{image}(\rho))\,[\rho(g(z)) = z].$$

(3) We say ρ is one-way *if and only if ρ is honest, polynomial-time computable, and noninvertible.*

We now define strong noninvertibility (or strongness, for short), which is a stand-alone property (i.e., a property with one-way-ness not necessarily required) of 2-ary functions. If one wants to discuss strongness in a nontrivial way, one needs some type of honesty that is suitable for strongness. To this end, we introduce below, in addition to honesty as defined above, the notion of s-honesty. [5]

Definition 2.2 (see, essentially, [RS97,HR99]) *Let $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ be any (possibly nontotal, possibly many-to-one) 2-ary function.*

(1) We say σ is s-honest *if and only if there exists a polynomial q such that both* (a) *and* (b) *hold:*
(a) $(\forall z, a : (\exists b)\,[\sigma(a,b) = z])\,(\exists b')\,[|b'| \leq q(|z| + |a|) \wedge \sigma(a,b') = z]$.
(b) $(\forall z, b : (\exists a)\,[\sigma(a,b) = z])\,(\exists a')\,[|a'| \leq q(|z| + |b|) \wedge \sigma(a',b) = z]$.

(2) We say σ is (polynomial-time) invertible with respect to the first argument *if and only if*

$$(\exists g_1 \in \mathrm{FP})\,(\forall z \in \mathrm{image}(\sigma))\,(\forall a, b : (a,b) \in \mathrm{domain}(\sigma) \wedge \sigma(a,b) = z)\,[\sigma(a, g_1(\langle a, z\rangle)) = z].$$

(3) We say σ is (polynomial-time) invertible with respect to the second argument *if and only if*

$$(\exists g_2 \in \mathrm{FP})\,(\forall z \in \mathrm{image}(\sigma))\,(\forall a, b : (a,b) \in \mathrm{domain}(\sigma) \wedge \sigma(a,b) = z)\,[\sigma(g_2(\langle b, z\rangle), b) = z].$$

(4) We say σ is strongly noninvertible *if and only if σ is neither invertible with respect to the first argument nor invertible with respect to the second argument.*

(5) We say σ is strongly one-way *if and only if σ is s-honest, polynomial-time computable, and strongly noninvertible.*

There are honest, polynomial-time computable 2-ary functions that are not s-honest, and there are s-honest, polynomial-time computable 2-ary functions that are not honest. As an example of the latter, consider the function $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ that is defined by $\sigma(a,b) = 1^{\lceil \log\log(\max(|a|,2))\rceil}$ if $|a| = |b|$, and that is undefined otherwise; this function is s-honest but not honest. As an example of the former, consider the function $\rho : \Sigma^* \times \Sigma^* \to \Sigma^*$ defined by $\rho(a,b) = 1^{\lceil \log\log(\max(|b|,2))\rceil}$ if $a = 0$, and $\rho(a,b) = ab$ if $a \neq 0$. This function is honest, as proven by $\rho(\epsilon, x) = x$. However, it is not s-honest, since for

[5] The strongly noninvertible functions in [HR99] are all s-honest, notwithstanding that s-honesty is not explicitly discussed in [HR99] (or in [RS97,RS93]).

any given polynomial q there are strings $b \in \Sigma^*$ and $z = 1^{\lceil \log\log(\max(|b|,2))\rceil}$ with $\rho(0,b) = z$, but the smallest string $b' \in \Sigma^*$ with $\rho(0,b') = z$ satisfies $|b'| > q(|z| + |0|) = q(\lceil \log\log(\max(|b|,2))\rceil + 1)$.

For completeness, we also give a formal definition of the notion of overstrongness mentioned in the introduction.

Definition 2.3 *Let $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ be any (possibly nontotal, possibly many-to-one) 2-ary function. We say σ is* overstrong *if and only if for no $f \in$ FP with $f : \{1,2\} \times \Sigma^* \times \Sigma^* \to \Sigma^* \times \Sigma^*$ does it hold that for each $i \in \{1,2\}$ and for all strings $z, a \in \Sigma^*$:*

$$\begin{aligned}&((\exists b \in \Sigma^*)[(\sigma(a,b) = z \wedge i = 1) \vee (\sigma(b,a) = z \wedge i = 2)])\\ &\implies \sigma(f(i,z,a)) = z.\end{aligned}$$

Note that overstrongness implies both noninvertibility and strong noninvertibility.

3 On Inverting Strongly Noninvertible Functions

It is well-known (see, e.g., [Sel92,BDG95]) that 1-ary one-way functions exist if and only if P $\neq$ NP. As mentioned in [HR99,RS97], the standard method to prove this result can also be used to prove the analogous result for 2-ary one-way functions.

Theorem 3.1 (see [HR99,RS97]) P $\neq$ NP *if and only if total 2-ary one-way functions exist.*

Now we show the main result of this paper: If P $\neq$ NP then one can invert some functions that are strongly noninvertible.

Theorem 3.2 *If* P $\neq$ NP *then there exists a total, honest 2-ary function that is a strongly one-way function but not a one-way function.*

Proof. Assuming P $\neq$ NP, by Theorem 3.1 there exists a total 2-ary one-way function ρ. Define a function $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ as follows:

$$\sigma(a,b) = \begin{cases} 0\rho(x,y) & \text{if } (\exists x,y,z \in \Sigma^*)\,[a = 1\langle x,y\rangle \wedge b = 0z] \\ 0\rho(y,z) & \text{if } (\exists x,y,z \in \Sigma^*)\,[a = 0x \wedge b = 1\langle y,z\rangle] \\ 1xy & \text{if } (\exists x,y \in \Sigma^*)\,[(a = 0x \wedge b = 0y) \vee (a = 1x \wedge b = 1y)] \\ ab & \text{if } a = \epsilon \vee b = \epsilon. \end{cases}$$

It is a matter of routine to check that σ is polynomial-time computable, total, honest, and (regardless of whether or not ρ, which is honest, is s-honest) s-honest.

If one could invert σ with respect to one of its arguments then one could invert ρ, contradicting that ρ is a one-way function. In particular, supposing σ is invertible with respect to the first argument via inverter $g_1 \in \mathrm{FP}$, we can use g_1 to define a function $g \in \mathrm{FP}$ that inverts ρ. To see this, note that given any $w \in \mathrm{image}(\rho)$ with $w \neq \epsilon$, the function g_1 on input $\langle 0, 0w \rangle$ must yield a string of the form $b = 1\langle y, z \rangle$ with $\rho(y, z) = w$. Thus, σ is not invertible with respect to the first argument. An analogous argument shows that σ is not invertible with respect to the second argument. It follows that σ is strongly noninvertible.

However, σ is invertible, since every string $z \in \mathrm{image}(\sigma)$ has an inverse of the form (ϵ, z); so, the FP function mapping any given string z to (ϵ, z) is an inverter for σ. Hence, σ is not a one-way function. ❑

The converse of Theorem 3.2 immediately holds, as do the converses of Proposition 3.3, Corollary 3.5, and Theorems 3.4, 3.6, and 3.7. However, although all these results in fact are equivalences, we will focus on only the interesting implication direction.

For completeness, we mention in passing that, assuming P $\neq$ NP, one can construct functions that—unlike the function constructed in the proof of Theorem 3.2—are simultaneously strongly one-way *and* one-way. An example of such a function is the following modification $\hat{\sigma}$ of the function σ constructed in the proof of Theorem 3.2. As in that proof, let ρ be a total 2-ary one-way function, and define function $\hat{\sigma} : \Sigma^* \times \Sigma^* \to \Sigma^*$ by

$$\hat{\sigma}(a, b) = \begin{cases} 0\rho(x, y) \text{ if } (\exists x, y, z \in \Sigma^*)\,[a = 1\langle x, y \rangle \wedge b = 0z] \\ 0\rho(y, z) \text{ if } (\exists x, y, z \in \Sigma^*)\,[a = 0x \wedge b = 1\langle y, z \rangle] \\ 1ab \qquad \text{otherwise.} \end{cases}$$

Note that $\hat{\sigma}$ even is overstrong; hence, $\hat{\sigma}$ is both noninvertible and strongly noninvertible. The following proposition captures this observation.

Proposition 3.3 *If* P $\neq$ NP *then there exists a total, honest, s-honest, 2-ary overstrong function. (It follows that if* P $\neq$ NP *then there exists a total 2-ary function that is one-way and strongly one-way.)*

Corollary 3.5 below shows that if P $\neq$ NP then there is an s-honest 2-ary one-way function that is not strongly one-way. First, we establish a result that is slightly stronger: For a function to be not strongly noninvertible, it is enough

that it is invertible with respect to at least one of its arguments. The function σ to be constructed in the proof of Theorem 3.4 below even is invertible with respect to each of its arguments.

Theorem 3.4 *If* P $\neq$ NP *then there exists a total, s-honest 2-ary one-way function σ such that σ is invertible with respect to its first argument and σ is invertible with respect to its second argument.*

Proof. It is well-known ([Sel92, Prop. 1], in light of the many-to-one analog of his comment [Sel92, p. 209] about totality) that under the assumption P $\neq$ NP there exists a total 1-ary one-way function $\rho : \Sigma^* \to \Sigma^*$. Define a function $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ as follows:

$$\sigma(a,b) = \begin{cases} 1\rho(a) & \text{if } a = b \\ 0ab & \text{if } a \neq b. \end{cases}$$

Note that σ is polynomial-time computable, total, s-honest, and honest. If σ were invertible in polynomial time then ρ would be too; so, σ is a one-way function. However, σ is invertible with respect to each of its arguments. For an inverter with respect to the first argument, consider the function $g_1 : \Sigma^* \to \Sigma^*$ defined by

$$g_1(x) = \begin{cases} b \text{ if } (\exists a, b, z \in \Sigma^*)\,[x = \langle a, 0z\rangle \wedge z = ab] \\ a \text{ if } (\exists a, z \in \Sigma^*)\,[x = \langle a, 1z\rangle] \\ \epsilon \text{ otherwise.} \end{cases}$$

Clearly, $g_1 \in$ FP. Note that for every $y \in \text{image}(\sigma)$ and for every $a \in \Sigma^*$ for which there exists some $b \in \Sigma^*$ with $\sigma(a,b) = y$, it holds that $\sigma(a, g_1(\langle a, y\rangle)) = y$, completing the proof that σ is invertible with respect to the first argument. To see that σ also is invertible with respect to the second argument, an analogous construction (with the roles of the first and the second argument interchanged) works to give an inverter g_2 for a fixed second argument. ❑

Theorem 3.4's construction has very recently been built on in work of Hemaspaandra, Rothe, and Saxena that shows that, for each of the 81 possible specifications (with respect to having, not having, or not caring about each of strongness, totality, commutativity, and associativity) of one-way functions, one-way functions of that type exist exactly if P and NP differ [HRS05]. In particular, Lemma 5.2 of that paper—which asserts that if P and NP differ then there exist total one-way functions that are not commutative, not associative, and not strongly noninvertible—is proven by directly invoking the above construction. And Lemma 5.1 of that paper—which asserts that if P and

NP differ then there exist total, commutative one-way functions that are neither associative nor strongly noninvertible—is proven by adapting the above construction to force commutativity.

Corollary 3.5 *If* P $\neq$ NP *then there exists a total, s-honest 2-ary one-way function that is not strongly one-way.*

One might wonder whether functions that are not strongly noninvertible (which means that they are invertible with respect to at least one of their arguments) outright must be invertible with respect to both of their arguments. The following result states that this is not the case in general, unless P = NP.

Theorem 3.6 *If* P $\neq$ NP *then there exists a total, s-honest 2-ary one-way function that is invertible with respect to one of its arguments (thus, it is not strongly one-way), yet that is not invertible with respect to its other argument.*

Proof. Assuming P $\neq$ NP, by Theorem 3.1 there exists a total 2-ary one-way function, call it ρ. Since our pairing function is onto and one-to-one, and its inverses are efficiently computable, the functions—π_1 and π_2—mapping from each string in Σ^* to that string's first and second components when interpreted as a pair are well-defined, total, polynomial-time functions; for all $b \in \Sigma^*$, $b = \langle \pi_1(b), \pi_2(b) \rangle$. Define a function $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ as follows:

$$\sigma(a, b) = \rho(\pi_1(b), \pi_2(b)).$$

It is clear that σ is honest (via ρ's honesty) and s-honest. Let a_0 be any fixed string, and define $g_2(w) = a_0$ for all strings w. Clearly, $g_2 \in$ FP. The definition of σ implies that for each $z = \rho(x, y) \in \text{image}(\sigma)$ and for each $b \in \Sigma^*$ such that $\sigma(a, b) = z$ for some $a \in \Sigma^*$, it also holds that $\sigma(a_0, b) = z$. Thus, σ is invertible with respect to the second argument via g_2. However, if σ were also invertible with respect to the first argument via some function $g_1 \in$ FP, then g_1 could be used to invert ρ, which would contradict the noninvertibility of ρ. Hence, σ is invertible with respect to its first, yet not with respect to its second argument. Analogously, we can define a function that is invertible with respect to its second argument, yet not with respect to its first argument. ❑

Finally, let us turn to the notion of overstrongness (see Definition 2.3) mentioned in the introduction. As noted there, this notion is not less restrictive than either noninvertibility or strong noninvertibility. That is, if a given polynomial-time computable, honest, s-honest function is overstrong then it certainly is both one-way and strongly one-way. As we alluded to in the introduction, overstrongness does not seem an appropriate notion to, even intuitively, underpin security within the cryptographic protocols of Rabi, Rivest, and Sherman. Nonetheless, for the purpose of showing that the notions do not

collapse, we will prove below that if P $\neq$ NP then overstrongness is a strictly more restrictive notion than both noninvertibility and strong noninvertibility.

Theorem 3.7 *If* P $\neq$ NP *then there exists a total, honest, s-honest 2-ary function that is noninvertible and strongly noninvertible but that is not overstrong.*

Proof. Assume P $\neq$ NP. It is known (see [Sel92]) that this assumption implies that total 1-ary one-way functions exist. Let $\widehat{\rho}$ be one such function, and let $\widehat{\rho}$ be such that it additionally satisfies

$$(\exists r \geq 2)\,(\forall x \in \Sigma^*)\,[|\widehat{\rho}(x)| = |x|^r + r]. \tag{3.1}$$

That condition (3.1) can be required follows easily from the standard "accepting-paths-based" proof that P $\neq$ NP implies the existence of total 1-ary one-way functions. Henceforth, r will denote one fixed value r satisfying condition (3.1).

Define a total function $\rho : \Sigma^* \to \Sigma^*$ as follows:

$$\rho(a) = \begin{cases} 1\widehat{\rho}(x) & \text{if } (\exists x \in \Sigma^*)\,[a = 1x] \\ a & \text{if } (\exists x \in \Sigma^*)\,[a = 0x] \\ \epsilon & \text{if } a = \epsilon. \end{cases}$$

Note that ρ is a 1-ary, total one-way function satisfying that for each $i \geq 0$, $\rho(0^i) = 0^i$. Now define the total function $\sigma : \Sigma^* \times \Sigma^* \to \Sigma^*$ as follows:

$$\sigma(a,b) = \begin{cases} 1\langle \rho(x), 0^{|y|} \rangle & \text{if } (\exists x, y \in \Sigma^*)\,[|x| = |y| \wedge a = 0\langle x, y\rangle = b] \\ 1\langle \rho(x), 0^{|y|} \rangle & \text{if } (\exists x, y \in \Sigma^*) \\ & \qquad [|x| = |y| \wedge a = 1\langle x, 0y\rangle \wedge b = 1\langle x, 1\widehat{\rho}(y)\rangle] \\ 1\langle \rho(x), 0^{|y|} \rangle & \text{if } (\exists x, y \in \Sigma^*) \\ & \qquad [|x| = |y| \wedge a = 1\langle x, 1\widehat{\rho}(y)\rangle \wedge b = 1\langle x, 0y\rangle] \\ 0\langle a, b\rangle & \text{otherwise.} \end{cases}$$

Clearly, σ is polynomial-time computable, honest, s-honest, and commutative. If σ were invertible, ρ would be too. Thus, σ is a one-way function.

Note that σ is strongly noninvertible, for if it could be inverted with respect to either argument then $\widehat{\rho}$ could be inverted too. Suppose, for example, σ were invertible with respect to the first argument via inverter $g_1 \in$ FP. Then $\widehat{\rho}$ could be inverted as follows. Given any $z \in \Sigma^*$, if there is no $k \in \mathbb{N}$ with

$k^r + r = |z|$, there is no inverse of z under $\widehat{\rho}$; so, in that case we may output anything. Otherwise (i.e., if there is a $k \in \mathbb{N}$ with $k^r + r = |z|$), run g_1 on input $\langle a, w\rangle$, where $a = 1\langle 0^k, 1z\rangle$ and $w = 1\langle 0^k, 0^k\rangle$. By the definition of σ, if $z \in \text{image}(\widehat{\rho})$, the result of $g_1(\langle a, w\rangle)$ must be of the form $1\langle 0^k, 0\hat{z}\rangle$ for some preimage $\hat{z}$ of z under $\widehat{\rho}$. Note that the equality $\widehat{\rho}(\hat{z}) = z$ can easily be verified, since $\widehat{\rho}$ is polynomial-time computable. A similar argument shows that σ is not invertible with respect to the second argument. Hence, σ is strongly one-way.

Finally, we claim that σ is not overstrong. Here is what an inverter f does when given $i = 1$, an alleged first argument $a \in \Sigma^*$ of σ, and an alleged output $z \in \Sigma^*$ of σ:

$$f(1, a, z) = \begin{cases} (x, y) & \text{if } (\exists x, y \in \Sigma^*)\,[z = 0\langle x, y\rangle] \\ (a, a) & \text{if } (\exists x, y \in \Sigma^*)\,(\exists m \in \mathbb{N}) \\ & \quad [a = 0x \wedge z = 1\langle y, 0^m\rangle] \\ (0\langle w, w\rangle, 0\langle w, w\rangle) & \text{if } (\exists w, x, y \in \Sigma^*)\,(\exists m \in \mathbb{N}) \\ & \quad [a = 1\langle w, 0x\rangle \wedge z = 1\langle y, 0^m\rangle] \\ (0\langle w, w\rangle, 0\langle w, w\rangle) & \text{if } (\exists w, x, y \in \Sigma^*)\,(\exists m \in \mathbb{N}) \\ & \quad [a = 1\langle w, 1x\rangle \wedge z = 1\langle y, 0^m\rangle] \\ (\epsilon, \epsilon) & \text{otherwise.} \end{cases}$$

Since σ is commutative, the above definition also shows how to handle the case $i = 2$.

Note that $f \in \text{FP}$. Whenever there exists some string $b \in \Sigma^*$ for which $\sigma(a, b) = z$, it holds that $\sigma(f(1, a, z)) = z$. (If there is no such b, it does not matter what $f(1, a, z)$ outputs.) Hence, σ is not overstrong. ❑

Speaking broadly, we would summarize the contribution of this paper as follows. One may view this paper as, at its core, a study of the naturalness and properties of the notion—which is the lynchpin of the security intuitions of the protocols of Rabi, Rivest, and Sherman—of "strong noninvertibility." Our results make it clear that the terminology itself was, in hindsight, poorly chosen, as our main result formally proves that strong noninvertibility is not strictly stronger than noninvertibility, under a very reasonable complexity-theoretic hypothesis, P $\neq$ NP. This result is not just about terminology, but rather tries to make clearer the degree of strength of the existing notion of strong noninvertibility, and to do so not via intuition but via a complexity-theoretic characterization (namely, in light of the comment immediately following Theorem 3.2, "exactly as likely as P $\neq$ NP").

Acknowledgments: We appreciate the helpful refereeing feedback and thank the editors. We are grateful to Chris Homan for suggesting that we formally define overstrongness. We thank Osamu Watanabe for mentioning to us the notions, different from those used here though slightly reminiscent, from average-case theory, of claw-free collections, collision-free pseudorandom generators, and collision-free hash functions.

References

[All86] E. Allender. The complexity of sparse sets in P. In *Proceedings of the 1st Structure in Complexity Theory Conference*, pages 1–11. Springer-Verlag *Lecture Notes in Computer Science #223*, June 1986.

[AR88] E. Allender and R. Rubinstein. P-printable sets. *SIAM Journal on Computing*, 17(6):1193–1202, 1988.

[BC93] D. Bovet and P. Crescenzi. *Introduction to the Theory of Complexity.* Prentice Hall, 1993.

[BDG95] J. Balcázar, J. Díaz, and J. Gabarró. *Structural Complexity I.* EATCS Monographs on Theoretical Computer Science. Springer-Verlag, second edition, 1995.

[Ber77] L. Berman. *Polynomial Reducibilities and Complete Sets.* PhD thesis, Cornell University, Ithaca, NY, 1977.

[BHHR99] A. Beygelzimer, L. Hemaspaandra, C. Homan, and J. Rothe. One-way functions in worst-case cryptography: Algebraic and security properties are on the house. *SIGACT News*, 30(4):25–40, December 1999.

[GS88] J. Grollmann and A. Selman. Complexity measures for public-key cryptosystems. *SIAM Journal on Computing*, 17(2):309–335, 1988.

[Hom04] C. Homan. Tight lower bounds on the ambiguity of strong, total, associative, one-way functions. *Journal of Computer and System Sciences*, 68(3):657–674, 2004.

[HR99] L. Hemaspaandra and J. Rothe. Creating strong, total, commutative, associative one-way functions from any one-way function in complexity theory. *Journal of Computer and System Sciences*, 58(3):648–659, 1999.

[HRS05] L. Hemaspaandra, J. Rothe, and A. Saxena. Enforcing and defying associativity, commutativity, totality, and strong noninvertibility for one-way functions in complexity theory. In *Proceedings of the Ninth Italian Conference on Theoretical Computer Science*, pages 265–279. Springer-Verlag *Lecture Notes in Computer Science #3701*, October 2005.

[Ko85] K. Ko. On some natural complete operators. *Theoretical Computer Science*, 37(1):1–30, 1985.

[KRS88] B. Kaliski Jr., R. Rivest, and A. Sherman. Is the data encryption standard a group? (Results of cycling experiments on DES). *Journal of Cryptology*, 1(1):3–36, 1988.

[Pap94] C. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1994.

[RS93] M. Rabi and A. Sherman. Associative one-way functions: A new paradigm for secret-key agreement and digital signatures. Technical Report CS-TR-3183/UMIACS-TR-93-124, Department of Computer Science, University of Maryland, College Park, Maryland, 1993.

[RS97] M. Rabi and A. Sherman. An observation on associative one-way functions in complexity theory. *Information Processing Letters*, 64(5):239–244, 1997.

[Sel92] A. Selman. A survey of one-way functions in complexity theory. *Mathematical Systems Theory*, 25(3):203–221, 1992.

[She86] A. Sherman. *Cryptology and VLSI (a Two-Part Dissertation).* PhD thesis, MIT, Cambridge, MA, 1986. Available as Technical Report MIT/LCS/TR-381.