Multi-Prover Commitments Against Non-Signaling Attacks

Serge Fehr and Max Fillinger

Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
{serge.fehr,M.J.Fillinger}@cwi.nl

Abstract

We reconsider the concept of two-prover (and more generally: multi-prover) commitments, as introduced in the late eighties in the seminal work by Ben-Or *et al.*. As was recently shown by Crépeau *et al.*, the security of known two-prover commitment schemes not only relies on the explicit assumption that the two provers cannot communicate, but also depends on what their information processing capabilities are. For instance, there exist schemes that are secure against classical provers but insecure if the provers have *quantum* information processing capabilities, and there are schemes that resist such quantum attacks but become insecure when considering general so-called *non-signaling* provers, which are restricted *solely* by the requirement that no communication takes place.

This poses the natural question whether there exists a two-prover commitment scheme that is secure under the *sole* assumption that no communication takes place, and that does not rely on any further restriction of the information processing capabilities of the dishonest provers; no such scheme is known.

In this work, we give strong evidence for a negative answer: we show that any single-round two-prover commitment scheme can be broken by a non-signaling attack. Our negative result is as bad as it can get: for any candidate scheme that is (almost) perfectly hiding, there exists a strategy that allows the dishonest provers to open a commitment to an arbitrary bit (almost) as successfully as the honest provers can open an honestly prepared commitment, i.e., with probability (almost) 1 in case of a perfectly sound scheme. In the case of multi-round schemes, our impossibility result is restricted to perfectly hiding schemes.

On the positive side, we show that the impossibility result can be circumvented by considering *three* provers instead: there exists a three-prover commitment scheme that is secure against arbitrary non-signaling attacks.

1 Introduction

Background. A commitment scheme is an important primitive in theoretical cryptography with various applications, for instance to zero-knowledge proofs and multiparty computation, which themselves are fundamentally important concepts in modern cryptography. For a commitment scheme to be secure, it must be *hiding* and *binding*. The former means that after the commit phase, the committed value is still hidden from the verifier, and the latter means that the prover (also referred to as committer) can open a commitment only to one value. Unfortunately, a commitment scheme cannot be unconditionally hiding *and* unconditionally binding at the same time. This is easy to see in the classical setting, and holds as well when using quantum communication [May97, LC97]. Thus, we have to put some limitation on the capabilities of the dishonest party. One common approach is to assume that the dishonest prover (or, alternatively, the dishonest verifier) has limited computing resources, so that he cannot solve certain computational problems (like factoring large integers). Another approach was suggested by Ben-Or, Goldwasser, Kilian and Wigderson in their seminal paper [BGKW88] in the late eighties. They assume that the prover consists of two (or more) agents that cannot communicate with each other, and they show the existence of a secure commitment scheme in this two-prover setting. Based on this two-prover commitment scheme, they then show that every language in NP has a two-prover perfect zero-knowledge interactive proof system (though there are some subtle issues in this latter result, as discussed in [Yan13]).

A simple example of a two-prover commitment scheme, due to [CSST11], is the following. The verifier chooses a uniformly random string $a \in \{0,1\}^n$ and sends it to the first prover, who sends back $x := r \oplus a \cdot b$ as the commitment for bit $b \in \{0,1\}$, where $r \in \{0,1\}^n$ is a uniformly random string known (only) to the two provers, and where "$\oplus$" is bit-wise XOR and "$\cdot$" scalar multiplication (of the scalar b with the

vector a). In order to open the commitment (to b), the second prover sends back $y := r$, and the verifier checks the obvious: whether $y = x \oplus a \cdot b$. It is clear that this scheme is hiding: $x := r \oplus a \cdot b$ is uniformly random and independent of a no matter what b is, and the intuition behind the binding property is the following. In order to open the commitment to $b = 0$, the second prover needs to announce $y = x$; in order to open to $b = 1$, he needs to announce $y = x \oplus a$. Therefore, in order to open to *both*, he must know x *and* $x \oplus a$, which means he knows a, but this is a contradiction to the no-communication assumption, because a was sent only to the first prover.

In [CSST11], Crépeau, Salvail, Simard and Tapp show that, as a matter of fact, the security of such two-prover commitment schemes not only relies on the explicit assumption that the two provers cannot communicate, but the security also crucially depends on the information processing capabilities of the dishonest provers. Indeed, they show that a slight variation of the above two-prover commitment scheme (where some slack is given to the verification $y = x \oplus a \cdot b$) is secure against classical provers, but is completely insecure if the provers have *quantum* information processing capabilities and can obtain x and y by means of doing local measurements on an entangled quantum state.[1] Furthermore, they show that the above example two-prover commitment scheme remains secure against such quantum attacks, but becomes insecure against so-called *non-signaling* provers. The notion of non-signaling was first introduced by Khalfin and Tsirelson [TK85] and by Rastall [Ras85] in the context of Bell-inequalities, and later reintroduced by Popescu and Rohrlich [PR94]. Non-signaling provers are restricted *solely* by the requirement that no communication takes place — no additional restriction limits their information processing capabilities (not even the laws of quantum mechanics) — and thus considering non-signaling provers is the *minimal* assumption for the two-prover setting to make sense.

This gives rise to the following question. Does there exist a two-prover commitment scheme that is secure against arbitrary non-signaling provers? Such a scheme would *truly* be based on the sole assumption that the provers cannot communicate. No such scheme is known. Clearly, from a practical point of view, asking for such a scheme may be overkill; given our strong believe in quantum mechanics, relying on a scheme that resists quantum attacks seems to be a safe bet. But from a theoretical perspective, this question is certainly in line with the general goal of theoretical cryptography: to find the strongest possible security based on the weakest possible assumption.

Our Results.   In this work, we give strong evidence for a negative answer: we show that there exists no single-round two-prover commitment scheme that is secure against general non-signaling attacks. Our impossibility result is as strong as it can get. We show that for any candidate single-round two-prover commitment scheme that is (almost) perfectly hiding, the binding property can be (almost) completely broken: there exists a non-signaling strategy that allows the dishonest provers to open a commitment to an arbitrary bit (almost) as successfully as the honest provers can open an honestly prepared commitment, i.e., with probability (almost) 1 in case of a perfectly sound scheme. Furthermore, for a restricted but natural class of schemes, namely for schemes that have the same communication pattern as the above example scheme, our impossibility result is tight: for every (rational) parameter $0 < \varepsilon \leq 1$ there exists a perfectly sound two-prover commitment scheme that is ε-hiding and as binding as allowed by our negative result (which is almost not binding if ε is small).

In the case of multi-round schemes, our impossibility result is limited and applies to perfectly hiding schemes only. Proving the impossibility of non-perfectly-hiding multi-round schemes remains open.

On the positive side, we show the existence of a secure *three*-prover commitment scheme against non-signaling attacks. Thus, our impossibility result can be circumvented by considering three instead of two provers.

Related Work.   Two-prover commitments are closely related to *relativistic commitments*, as introduced by Kent in [Ken99]. In a nutshell, a relativistic commitment scheme is a two-prover commitment scheme where the no-communication requirement is enforced by having the actions of the two provers separated by a space-like interval, i.e., the provers are placed far enough apart, and the scheme is executed quickly enough, so that no communication can take place by the laws of special relativity. As such, our impossibility result immediately implies impossibility of relativistic commitment schemes of the form we consider (e.g., we do not consider quantum schemes) against general non-signaling attacks.

Very generally speaking, and somewhat surprisingly, the (in)security of cryptographic primitives against non-signaling attacks may have an impact on more standard cryptographic settings, as was recently demonstrated by Kalai, Raz and Rothblum [KRR14], who showed the (computational) security

---

[1] The above intuition for the binding property of the scheme (which also applies to the variation considered in [CSST11]) fails in the quantum setting where x and y are obtained by means of *destructive* measurements.

of a *delegation scheme* based on the security of an underlying multi-party interactive proof system against non-signaling (or statistically-close-to-non-signaling) adversaries.

2 Preliminaries

2.1 (Conditional) Distributions

For the purpose of this work, a *(probability) distribution* is a function $p : \mathcal{X} \to \mathbb{R}$, $x \mapsto p(x)$, where $\mathcal{X}$ is a finite non-empty set, with the properties that $p(x) \geq 0$ for every $x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} p(x) = 1$. For any subset $\Lambda \subset \mathcal{X}$, $p(\Lambda)$ is naturally defined as $p(\Lambda) = \sum_{x \in \Lambda} p(x)$, and it holds that

$$p(\Lambda) + p(\Gamma) = p(\Lambda \cup \Gamma) - p(\Lambda \cap \Gamma) \leq 1 + p(\Lambda \cap \Gamma) \tag{1}$$

for all $\Lambda, \Gamma \subset \mathcal{X}$. A probability distribution is *bipartite* if it is of the form $p : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. In case of such a bipartite distribution $p(x, y)$, probabilities like $p(x\!=\!y)$, $p(x\!=\!f(y))$, $p(x \neq y)$ etc. are naturally understood as

$$p(x\!=\!y) = p(\{(x, y) \in \mathcal{X} \times \mathcal{Y} \,|\, x = y\}) = \sum_{\substack{x \in \mathcal{X}, y \in \mathcal{Y} \\ \text{s.t. } x = y}} p(x, y)$$

etc. Also, for a bipartite distribution $p : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, the *marginals* $p(x)$ and $p(y)$ are given by $p(x) = \sum_y p(x, y)$ and $p(y) = \sum_x p(x, y)$, respectively. We note that this notation may lead to an ambiguity when writing $p(w)$ for some $w \in \mathcal{X} \cap \mathcal{Y}$; we avoid this by writing $p(x\!=\!w)$ or $p(y\!=\!w)$ instead, which are naturally understood. The above obviously extends to arbitrary *multipartite* distributions $p(x, y, z)$ etc.

A *conditional (probability) distribution* is a function $p : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$, $(x, a) \mapsto p(x|a)$, for finite non-empty sets $\mathcal{X}$ and $\mathcal{A}$, such that for every fixed $a^* \in \mathcal{A}$, the function $p(x|a^*)$ is a probability distribution in the above sense, which we also write as $p(x|a\!=\!a^*)$. As such, the above naturally extends to bi- and multipartite conditional probability distributions; e.g., if $p(x, y|a, b)$ is a conditional distribution then $p(x|a, b)$, $p(y|a, b)$, $p(x\!=\!y|a, b)$ etc. are all naturally defined. However, we emphasize that for instance $p(x|a)$ is in general *not* well defined — unless the corresponding conditional distribution $p(b|a)$ is given, or unless $p(x|a, b)$ does not depend on b.

Remark 2.1. *By convention, we write $p(x|a, b) = p(x|a)$ to express that $p(x|a, b)$ does not depend on b, i.e., that $p(x|a, b_1) = p(x|a, b_2)$ for all b_1 and b_2, and as such $p(x|a)$ is well defined and equals $p(x|a, b)$.*

A distribution $\delta(x)$ over $\mathcal{X}$ is called a *Dirac* distribution if there exists $x^* \in \mathcal{X}$ so that $\delta(x\!=\!x^*) = 1$, and a conditional distribution $\delta(x|a)$ over $\mathcal{X}$ is called a conditional *Dirac* distribution if $\delta(x|a\!=\!a^*)$ is a *Dirac* distribution for every $a^* \in \mathcal{A}$, i.e., for every $a^* \in \mathcal{A}$ there exists $x^* \in \mathcal{X}$ so that $\delta(x\!=\!x^*|a\!=\!a^*) = 1$.

Note that we often abuse notation slightly and simply write $p(x)$ instead of $p : \mathcal{X} \to \mathbb{R}$, $x \mapsto p(x)$; furthermore, we may use p for different distributions and distinguish between them by using different names for the variable, like when we consider the two marginals $p(x)$ and $p(y)$ of a bipartite distribution $p(x, y)$. Finally, given two distributions $p(x_0)$ and $q(x_1)$ over the same set $\mathcal{X}$ (and similarly if we use the above convention and denote them by $p(x_0)$ and $p(x_1)$ instead), we write $p(x_0) = q(x_1)$ to denote that $p(x_0 = w) = q(x_1 = w)$ for all $w \in \mathcal{X}$. In a corresponding way, equalities like $p(x_0, x_0', y) = q(x_1, x_1', y)$ should be understood; in situations where we feel it is helpful, we may clarify that "x_0 is associated with x_1, and x_0' with x_1'"; similarly for conditional distributions.

2.2 Gluing Together Distributions

We recall the definition of the statistical distance.

Definition 2.2. *Let $p(x_0)$ and $p(x_1)$ be two distributions over the same set $\mathcal{X}$.[2] Then, their statistical distance is defined as*

$$d\big(p(x_0), p(x_1)\big) = \frac{1}{2} \cdot \sum_{x \in \mathcal{X}} \big|p(x_0\!=\!x) - p(x_1\!=\!x)\big|.$$

The following property of the statistical distance is well known (see e.g. [RK05]).

---

[2]This is without loss of generality: the domain can always be extended by including zero-probability elements.

Proposition 2.3. *Let $p(x_0)$ and $p(x_1)$ be two distributions over the same set $\mathcal{X}$ with $d\big(p(x_0), p(x_1)\big) = \varepsilon$. Then, there exists a distribution $p'(x_0, x_1)$ over $\mathcal{X} \times \mathcal{X}$ with marginals $p'(x_0) = p(x_0)$ and $p'(x_1) = p(x_1)$, and such that $p'(x_0 \neq x_1) = \varepsilon$.*

The following is an immediate consequence.

Lemma 2.4. *Let $p(x_0, y_0)$ and $p(x_1, y_1)$ be distributions with $d\big(p(x_0), p(x_1)\big) = \varepsilon$. Then, there exists a distribution $p'(x_0, x_1, y_0, y_1)$ with marginals $p'(x_0, y_0) = p(x_0, y_0)$ and $p'(x_1, y_1) = p(x_1, y_1)$, and such that $p'(x_0 \neq x_1) = \varepsilon$ and, as a consequence, $d\big(p'(x_0, y_1), p'(x_1, y_1)\big) \leq \varepsilon$.*

Proof. We first apply Proposition 2.3 to $p(x_0)$ and $p(x_1)$ to obtain $p'(x_0, x_1)$, and then we set

$$p'(x_0, x_1, y_0, y_1) = p'(x_0, x_1) \cdot p(y_0|x_0) \cdot p(y_1|x_1).$$

The claims on the marginals and on $p'(x_0 \neq x_1)$ follow immediately, and for the last claim we note that

$$\begin{aligned}
p'(x_0, y_1) &= p'(x_0 = x_1) \cdot p'(x_0, y_1|x_0 = x_1) + p'(x_0 \neq x_1) \cdot p'(x_0, y_1|x_0 \neq x_1) \\
&= p'(x_0 = x_1) \cdot p'(x_1, y_1|x_0 = x_1) + p'(x_0 \neq x_1) \cdot p'(x_0, y_1|x_0 \neq x_1)
\end{aligned}$$

and

$$p'(x_1, y_1) = p'(x_0 = x_1) \cdot p'(x_1, y_1|x_0 = x_1) + p'(x_0 \neq x_1) \cdot p'(x_1, y_1|x_0 \neq x_1)$$

and the claim follows because $p'(x_0 \neq x_1) = \varepsilon$. $\qquad\square$

Remark 2.5. *Note that due to the consistency of the marginals, it makes sense to write $p(x_0, x_1, y_0, y_1)$ instead of $p'(x_0, x_1, y_0, y_1)$. We say that we "glue together" $p(x_0, y_0)$ and $p(x_1, y_1)$ along x_0 and x_1.*

Remark 2.6. *In the special case where $p(x_0)$ and $p(x_1)$ are identically distributed, i.e., $d\big(p(x_0), p(x_1)\big) = 0$, we obviously have $p(x_0, y_1) = p(x_1, y_1)$.*

Remark 2.7. *It is easy to see from the proof of Lemma 2.4 that the following natural property holds. If $p(x_0, x_1, y_0, y_1, y_0', y_1')$ is obtained by gluing together $p(x_0, y_0, y_0')$ and $p(x_1, y_1, y_1')$ along x_0 and x_1, then the marginal $p(x_0, x_1, y_0, y_1)$ coincides with the distribution obtained by gluing together the marginals $p(x_0, y_0)$ and $p(x_1, y_1)$ along x_0 and x_1.*

3 Bipartite Systems and Two-Prover Commitments

3.1 One-Round Bipartite Systems

Informally, a *bipartite system* consists of two subsystem, which we refer to as the left and the right subsystem. Upon input a to the left and input a' to the right subsystem, the left subsystem outputs x and the right subsystem outputs x' (see Figure 1, left). Formally, the behavior of such a system is given by a conditional distribution $q(x, x'|a, a')$, with the interpretation that upon input pair (a, a'), the system outputs a specific pair (x, x') with probability $q(x, x'|a, a')$. Note that we leave the sets $\mathcal{A}, \mathcal{A}', \mathcal{X}$ and $\mathcal{X}'$, from which a, a', x and x' are respectively sampled, implicit.

If we do not put any restriction upon the system, then *any* conditional distribution $q(x, x'|a, a')$ is eligible, i.e., describes a bipartite system. However, we are interested in systems where the two subsystems cannot communicate with each other. How exactly this requirement restricts $q(x, x'|a, a')$ depends on the available "resources". For instance, if the two subsystems are deterministic, i.e., compute x and x' as *deterministic* functions of a and a' respectively, then this restricts $q(x, x'|a, a')$ to be of the form $q(x, x'|a, a') = \delta(x|a) \cdot \delta(x'|a')$ for conditional Dirac distributions $\delta(x|a)$ and $\delta(x'|a')$. If in addition to allowing them to compute deterministic functions, we give the two subsystem *shared randomness*, then $q(x, x'|a, a')$ may be of the form

$$q(x, x'|a, a') = \sum_r p(r) \cdot \delta(x|a, r) \cdot \delta(x'|a', r)$$

for a distribution $p(r)$ and conditional Dirac distributions $\delta(x|a, r)$ and $\delta(x'|a', r)$. Such a system is called *classical* or *local*. Interestingly, this is not the end of the story. By the laws of *quantum mechanics*, if the two subsystems share an entangled quantum state and obtain x and x' without communication

as the result of local measurements that may depend on a and a', respectively, then this gives rise to conditional distributions $q(x, x'|a, a')$ of the form

$$q(x, x'|a, a') = \langle \psi | \big(E_x^a \otimes F_{x'}^{a'} \big) | \psi \rangle \,,$$

where $|\psi\rangle$ is a quantum state and $\{E_x^a\}_x$ and $\{F_{x'}^{a'}\}_{x'}$ are so-called POVMs. What this exactly means is not important for us; what *is* important is that this leads to a *strictly larger* class of bipartite systems. This is typically referred to as a *violation of Bell inequalities* [Bel64], and is nicely captured by the notion of *nonlocal games*. A famous example is the so-called CHSH-game [CHSH69], which is closely connected to the example two-prover commitment scheme from the introduction, and which shows that the variant considered in [CSST11] is insecure against quantum attacks.

The largest possible class of bipartite systems that is compatible with the requirement that the two subsystem do not communicate, but otherwise does not assume anything on the available resources and/or the underlying physical theory, are the so-called *non-signaling* systems, defined as follows.

Definition 3.1. *A conditional distribution $q(x, x'|a, a')$ is called a* non-signaling (one-round) bipartite system *if it satisfies*

$$q(x|a, a') = q(x|a) \qquad \text{(NS)}$$

as well as with the roles of the primed and unprimed variables exchanged, i.e.,

$$q(x'|a, a') = q(x'|a') \qquad \text{(NS')}$$

Recall that, by the convention in Remark 2.1, the equality (NS) is to be understood in the sense that $q(x|a, a')$ does not depend on a', i.e., that $q(x|a, a'_1) = q(x|a, a'_2)$ for all a'_1, a'_2, and correspondingly for (NS').

We emphasize that this is the *minimal* necessary condition for the requirement that the two subsystems do not communicate. Indeed, if e.g. $q(x|a, a'_1) \neq q(x|a, a'_2)$, i.e., if the input-output behavior of the left subsystem depends on the input to the right subsystem, then the system can be used to communicate by giving input a'_1 or a'_2 to the right subsystem, and observing the input-output behavior of the left subsystem. Thus, in such a system, communication does take place.

The non-signaling requirement for a bipartite system is — conceptually and formally — equivalent to requiring that the two subsystems can (in principle) be queried *in any order*. Conceptually, it holds because the left subsystem should be able to deliver its outputs *before* the right subsystem has received any input if and only if the output does not depend on the right subsystem's input (which means that no information is communicated from right to left), and similarly the other way round. And, formally, we see that the non-signaling requirement from Definition 3.1 is equivalent to asking that $q(x, x'|a, a')$ can be written as

$$q(x, x'|a, a') = q(x|a) \cdot q(x'|x, a, a') \quad \text{and} \quad q(x, x'|a, a') = q(x'|a') \cdot q(x|x', a, a')$$

for some respective conditional distributions $q(x|a)$ and $q(x'|a')$. This characterization is a convenient way to "test" whether a given bipartite system is non-signaling without doing the maths.

Clearly, all classical systems are non-signaling. Also, any quantum system is non-signaling.[3] But there are non-signaling systems that are not quantum (and thus in particular not classical). The typical example is the *NL-box* (non-local box; also known as *PR-box*) [PR94], which, upon input bits a and a' outputs *random* output bits x and x' subject to

$$x \oplus x' = a \cdot a' \,.$$

This system is indeed non-signaling, as it can be queried in any order: submit a to the left subsystem to obtain a uniformly random x, and then submit a' to the right subsystem to obtain $x' := x \oplus a \cdot b$, and correspondingly the other way round.

3.2 Two-Round Systems

We now consider bipartite systems as discussed above, but where one can interact with the two subsystems multiple times. We restrict to two rounds: after having input a to the left subsystem and obtained x as output, one can now input b into the left subsystem and obtain output y, and similarly with the right subsystem (see Figure 1, right). In such a two-round setting, the non-signaling condition needs to
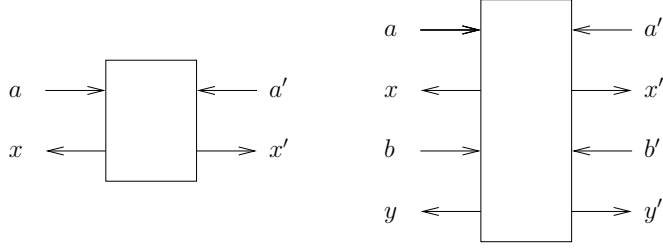
Figure 1: A one-round (left) and two-round (right) bipartite system.

be paired with *causality*, which captures that the output of the first round does not depend on the input that will be given in the second round.

Definition 3.2. *A conditional distribution* $q(x, x', y, y'|a, a', b, b')$ *is called a* non-signaling two-round bipartite system *if it satisfies the following two causality constraints*

$$q(x, x'|a, a', b, b') = q(x, x'|a, a') \tag{C1}$$
$$and \quad q(x'|x, y, a, a', b, b') = q(x'|x, y, a, a', b) \tag{C2}$$

and the following two non-signaling constraints

$$q(x, y|a, a', b, b') = q(x, y|a, b) \tag{NS1}$$
$$and \quad q(y|x, x', a, a', b, b') = q(y|x, x', a, a', b) \tag{NS2}$$

as well as with the roles of the primed and unprimed variables exchanged.

(C1) captures causality of the overall system, i.e., when considering the left and the right system as one "big" multi-round system. (C2) captures that no matter what interaction there is with the left system, the right system still satisfies causality. Similarly, (NS1) captures that the left and the right system are non-signaling over both rounds, and (NS2) captures that no matter what interaction there was in the first round, the left and the right system remain non-signaling in the second round.

It is rather clear that these are *necessary* conditions; we argue that they are *sufficient* to capture a non-signaling two-round system in Appendix A.

3.3 Two-Prover Commitments

We consider two-prover commitments of the following form. To commit to bit b, the two provers P and Q receive respective "questions" a and a' from the verifier V, and they compute, without communicating with each other, respective replies x and x' and send them to V. To open the commitment, P and Q send respectively y and y'. Finally, V performs some check to decide whether to accept or not.

In case of *classical* provers P and Q, restricting the opening phase to one round with one-way communication is without loss of generality: one may always assume that in the opening phase P and Q simply reveal the shared randomness, and V checks whether x and x' had been correctly computed, consistent with the claimed bit b. Restricting the commit phase to one round is, as far as we can see, *not* without loss of generality; we discuss the multi-round case later.

Formally, this can be captured as follows.

Definition 3.3. *A (single-round) two-prover commitment scheme* Com *consists of a probability distribution* $p(a, a')$, *two conditional distributions* $p_0(x, x', y, y'|a, a')$ *and* $p_1(x, x', y, y'|a, a')$, *and an acceptance predicate* $\mathsf{Acc}(x, x', y, y'|a, a', b)$.
We say that Com *is classical/quantum/non-signaling if* $p_0(x, x', y, y'|a, a')$ *and* $p_1(x, x', y, y'|a, a')$ *are both classical/quantum/non-signaling when parsed as bipartite one-round systems* $p_b((x, y), (x', y')|a, a')$. *By default, any two-prover commitment scheme* Com *is assumed to be non-signaling.*

The distribution $p(a, a')$ captures how V samples the "questions" a and a', $p_b(x, x', y, y'|a, a')$ describes the choices of x and x' and of y and y', given that the bit to commit to is b, and $\mathsf{Acc}(x, x', y, y'|a, a', b)$

---

[3]Indeed, the two parts of an entangled quantum state can be measured in any order, and the outcome of the first measurement does not depend on how the other part is going to be measured.

determines whether V accepts the opening or not. Whether a scheme is classical, quantum or non-signaling captures the restrictions of the honest provers.

Given a two-prover commitment scheme Com, we define

$$\text{Prob}[\text{Acc}|b] := \sum_{a,a',x,x',y,y'} p(a,a') \cdot p_b(x,x',y,y'|a,a') \cdot \text{Acc}(x,x',y,y'|a,a',b),$$

which is the probability that a correctly formed commitment to bit b is successfully opened.

Definition 3.4. *A commitment scheme* Com *is* θ-*sound if* $\text{Prob}_p[\text{Acc}|b] \geq \theta$ *for* $b \in \{0,1\}$. *We say that it is* perfectly sound *if it is 1-sound.*

It will be convenient to write $p(x_0, x_0', y_0, y_0'|a,a')$ instead of $p_0(x,x',y,y'|a,a')$ and $p(x_1,x_1',y_1,y_1'|a,a')$ instead of $p_1(x,x',y,y'|a,a')$. Switching to this notation, the hiding property is expressed as follows.

Definition 3.5. Com *is called* ε-*hiding if* $d\big(p(x_0,x_0'|a,a'),p(x_1,x_1'|a,a')\big) \leq \varepsilon$ *for all* a,a'. *If* Com *is* 0-*hiding, we also say it is* perfectly hiding.

Capturing the binding property is more subtle. From the classical approach of defining the binding property for a commitment scheme, one is tempted to require that once the commit phase is over and a, a', x and x' are fixed, adversarial provers $\hat{P}$ and $\hat{Q}$ cannot come up with an opening to $b = 0$ and *simultaneously* with an opening to $b = 1$, i.e., with y_0, y_0' and y_1, y_1' such that $\text{Acc}(x,x',y_0,y_0'|a,a',b{=}0)$ and $\text{Acc}(x,x',y_1,y_1'|a,a',b{=}1)$ are both satisfied (except with small probability). However, as pointed out by Dumais, Mayers and Salvail [DMS00], in the context of a general physical theory where y and y' may possibly be obtained as respective outcomes of *destructive* measurements (as is the case in quantum mechanics), such a definition is too weak. It does not exclude that $\hat{P}$ and $\hat{Q}$ can freely choose to open the commitment to $b = 0$ or to $b = 1$, whatever they want, but they cannot do both *simultaneously*; once they have produced one opening, their respective states got disturbed and the other opening can then not be obtained anymore.

Our definition for the binding property is based on the following game between the (honest) verifier V and the adversarial provers $\hat{P}, \hat{Q}$.

1. The commit phase is executed: V samples a and a' according to $p(a,a')$, and sends a to $\hat{P}$ and a' to $\hat{Q}$, upon which $\hat{P}$ and $\hat{Q}$ send x and x' back to V, respectively.

2. V sends a bit $b \in \{0,1\}$ to $\hat{P}$ and $\hat{Q}$.

3. $\hat{P}$ and $\hat{Q}$ try to open the commitment to b: they prepare y and y' and send them to V.

4. V checks if the verification predicate $\text{Acc}(x,x',y,y'|a,a',b)$ is satisfied.

We emphasize that even though in the actual binding game above, *the same* bit b is given to the two provers, we require that the response of the provers is well determined by their strategy even in the case that $b \neq b'$. Of course, if the provers are allowed to communicate, they are able to detect when $b \neq b'$ and could reply with, e.g., $y = y' = \perp$ in that case. However, if we restrict to non-signaling provers, we assume that it is *physically* impossible for them to communicate with each other and distinguish the case of $b = b'$ from $b \neq b'$.

As such, a non-signaling attack strategy against the binding property of a two-prover commitment scheme Com is given by a non-signaling two-round bipartite system $q(x,x',y,y'|a,a',b,b')$, as specified in Definition 3.2. For any such bipartite system, representing a strategy for $\hat{P}$ and $\hat{Q}$ in the above game, the probability that $\hat{P}$ and $\hat{Q}$ win the game, in that $\text{Acc}(x,x',y,y'|a,a',b)$ is satisfied when they have to open to the bit b, is given by

$$\text{Prob}_q^*[\text{Acc}|b] := \sum_{a,a',x,x',y,y'} p(a,a') \cdot q(x,x',y,y'|a,a',b,b) \cdot \text{Acc}(x,x',y,y'|a,a',b).$$

We are now ready to define the binding property.

Definition 3.6. *A two-prover commitment scheme* Com *is* δ-*binding (against non-signaling attacks) if it holds for any non-signaling two-round bipartite system* $q(x,x',y,y'|a,a',b,b')$ *that*

$$\text{Prob}_q^*[\text{Acc}|0] + \text{Prob}_q^*[\text{Acc}|1] \leq 1 + \delta.$$

In other words, a scheme is δ-binding if in the above game the dishonest provers win with probability at most $(1 + \delta)/2$ when $b \in \{0, 1\}$ is chosen uniformly at random. If a commitment scheme is binding (for a small δ) in the sense of Definition 3.6, then for any strategy q for $\hat{P}$ and $\hat{Q}$, they can just as well *honestly* commit to a bit $\hat{b}$, where $\hat{b}$ is set to 0 with probability $p_0 = \text{Prob}^*_q[\text{Acc}|0]$ and to 1 with probability $p_1 = 1 - p_0 \approx \text{Prob}^*_q[\text{Acc}|1]$, and they will have essentially the same respective success probabilities in opening the commitment to $b = 0$ and to $b = 1$.

4 Impossibility of Two-Prover Commitments

In this section, we show impossibility of secure single-round two-prover commitments against arbitrary non-signaling attacks. We start with the analysis of a restricted class of schemes which are easier to understand and for which we obtained stronger results.

4.1 Simple Schemes

We first consider a special, yet natural, class of schemes. We call a two-prover commitment scheme Com *simple* if it has the same communication pattern as the scheme described in the introduction. More formally, it is called simple if a', x' and y are "empty" (or fixed), i.e., if Com is given by $p(a)$, $p_0(x, y'|a)$, $p_1(x, y'|a)$ and $\text{Acc}(x, y'|a, b)$; to simplify notation, we then write y instead of y'. In other words, P is only involved in the commit phase, where, in order to commit to bit b, he outputs x upon input a, and Q is only involved in the opening phase, where he outputs y. The non-signaling requirement for Com then simplifies to $p_b(y|a) = p_b(y)$. Recall that by our convention, we may write $p(x_0, y_0|a)$ instead of $p_0(x, y|a)$ and $p(x_1, y_1|a)$ instead of $p_1(x, y|a)$.

In case of such a simple two-prover commitment scheme Com, a non-signaling two-prover strategy reduces to a non-signaling *one-round* bipartite system as specified in Definition 3.1 (see Figure 2).
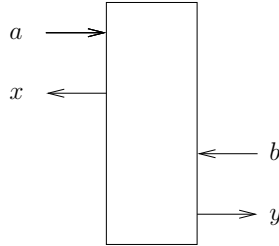

Figure 2: The adversaries' strategy $q(x, y|a, b)$ in case of a *simple* commitment scheme.

As a warm-up exercise, we first consider a simple two-prover commitment scheme that is *perfectly hiding* and *perfectly sound*. Recall that formally, a simple scheme is given by $p(a)$, $p_0(x, y|b)$, $p_1(x, y|a)$ and $\text{Acc}(x, y|a, b)$, and the perfect hiding property means that $p_0(x|a) = p_1(x|a)$ for any a. To show that such a scheme cannot be binding, we have to show that there exists a non-signaling one-round bipartite system $q(x, y|a, b)$ such that $\text{Prob}^*_q[\text{Acc}|0] + \text{Prob}^*_q[\text{Acc}|1]$ is significantly larger than 1. But this is actually trivial: we can simply set $q(x, y|a, b) := p_b(x, y|a)$. It then holds trivially that

$$\text{Prob}^*_q[\text{Acc}|b] = \sum_{a,x,y} p(a) \, q(x, y|a, b) \, \text{Acc}(x, y|a, b)$$
$$= \sum_{a,x,y} p(a) \, p_b(x, y|a) \, \text{Acc}(x, y|a, b)$$
$$= \text{Prob}_p[\text{Acc}|b]$$

and thus that the dishonest provers are as successful in opening the commitment as are the honest provers in opening an honestly prepared commitment. Thus, the binding property is broken as badly as it can get. The only thing that needs to be verified is that $q(x, y|a, b)$ is non-signaling, i.e., that $q(x|a, b) = q(x|a)$ and $q(y|a, b) = q(y|b)$. To see that the latter holds, note that $q(y|a, b) = p_b(y|a)$, and because Com is non-signaling we have that $p_b(y|a) = p_b(y)$, i.e., does not depend on a. Thus, the same holds for $q(y|a, b)$ and we have $q(y|a, b) = q(y|b)$. The former condition follows from the (perfect) hiding property: $q(x|a, b) = p_b(x|a) = p_{b'}(x|a) = q(x|a, b')$ for arbitrary $b, b' \in \{0, 1\}$, and thus $q(x|a, b) = q(x|a)$.

Below, we show how to extend this result to non-perfectly-binding simple schemes. In this case, we cannot simply set $q(x, y|a, b) := p_b(x, y|a)$, because such a q would not be non-signaling anymore — it would merely be "almost non-signaling". Instead, we have to find a strategy $q(x, y|a, b)$ that is (perfectly) non-signaling and close to $p_b(x, y|a)$; we will find such a strategy with the help of Lemma 2.4. In Section 4.2, we will then consider general schemes where *both* provers interact with the verifier in *both* phases. In this general case, further complications arise.

Theorem 4.1. *Consider a simple two-prover commitment scheme* Com *that is ε-hiding. Then, there exists a non-signaling strategy $q(x, y|a, b)$ such that*

$$\mathrm{Prob}_q^*[\mathsf{Acc}|0] = \mathrm{Prob}_p[\mathsf{Acc}|0] \quad and \quad \mathrm{Prob}_q^*[\mathsf{Acc}|1] \geq \mathrm{Prob}_p[\mathsf{Acc}|1] - \varepsilon \,.$$

If Com *is perfectly sound, it follows that*

$$\mathrm{Prob}_q^*[\mathsf{Acc}|0] + \mathrm{Prob}_q^*[\mathsf{Acc}|1] \geq 1 + (1 - \varepsilon)$$

and thus it cannot be δ-binding for $\delta < 1 - \varepsilon$.

Proof. Recall that Com is given by $p(a)$, $p_b(x, y|a)$ and $\mathsf{Acc}(x, y|a, b)$, and we write $p(x_b, y_b|a)$ instead of $p_b(x, y|a)$. Because Com is ε-hiding, it holds that $d\big(p(x_0|a), p(x_1|a)\big) \leq \varepsilon$ for any fixed a. Thus, using Lemma 2.4 for every a, we can glue together $p(x_0, y_0|a)$ and $p(x_1, y_1|a)$ along x_0 and x_1 to obtain a distribution $p(x_0, x_1, y_0, y_1|a)$ such that $p(x_0 \neq x_1|a) \leq \varepsilon$, and in particular $d\big(p(x_0, y_1|a), p(x_1, y_1|a)\big) \leq \varepsilon$.

We define a strategy q for the dishonest provers by setting $q(x, y|a, b) := p(x_0, y_b|a)$ (see Figure 3). First, we show that q is non-signaling. Indeed, we have $q(x|a, b) = p(x_0|a)$ for any b, so $q(x|a, b) = q(x|a)$, and we have $q(y|a, b) = p(y_b|a) = p(y_b)$ for any a, and thus $q(y|a, b) = q(y|b)$.

As for the acceptance probability, for $b = 0$ we have $q(x, y|a, 0) = p(x_0, y_0|a)$ and as such $\mathrm{Prob}_q^*[\mathsf{Acc}|0]$ equals $\mathrm{Prob}_p[\mathsf{Acc}|0]$. For $b = 1$, we have

$$d\big(q(x, y|a, 1), p(x_1, y_1|a)\big) = d\big(p(x_0, y_1|a), p(x_1, y_1|a)\big) \leq \varepsilon$$

and since the statistical distance does not increase under data processing, it follows that $\mathrm{Prob}_p[\mathsf{Acc}|1]$ and $\mathrm{Prob}_q^*[\mathsf{Acc}|1]$ are ε-close; this proves the claim. $\square$
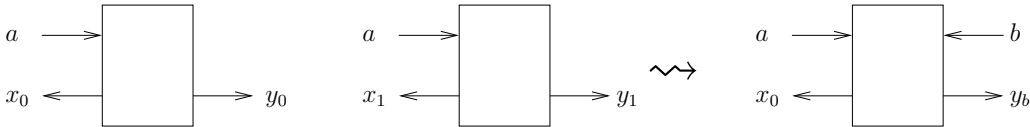
Figure 3: Defining the strategy q by gluing together $p(x_0, y_0|a)$ and $p(x_1, y_1|a)$.

The bound on the binding property in Theorem 4.1 is tight, as the following theorem shows.

Theorem 4.2. *For all $\varepsilon \in \mathbb{Q}$ such that $0 < \varepsilon \leq 1$ there exists a classical simple two-prover commitment scheme that is perfectly sound, ε-hiding and $(1 - \varepsilon)$-binding against non-signaling adversaries.*

Proof. We construct a scheme where the first prover reveals the bit b right at the beginning with probability ε. For simplicity, we first assume that $\varepsilon = 1/n$ for some integer $n \geq 1$ and then indicate how to extend the proof to arbitrary rational numbers.

The scheme works as follows. Let $[n] = \{0, \ldots, n - 1\}$. The shared randomness of the provers is $r \in [n]$ selected uniformly at random. The verifier selects $a \in [n]$ uniformly at random and sends it to prover P. If $a = r$ then P reveals $x := b$ to the verifier. Otherwise, he sends back $x := \bot$. In the opening phase, Q sends r to the verifier. The verifier accepts if and only if P revealed b or the output y of Q satisfies $y \in [n]$ and $y \neq a$.

It is clear that this scheme is sound and ε-hiding. Now consider dishonest provers that follow some non-signaling strategy $q(x, y|a, b)$. This then defines $q(a, x, y|b) = p(a) q(x, y|a, b)$ with $p(a) = 1/n$, and it holds that $\mathrm{Prob}_q^*[\mathsf{Acc}|b] = q(x = b|b) + q(x = \bot, y \neq a|b)$. Since $q(y|a, b) = q(y|b)$, we have

$$q(y \neq a|b) = \sum_{\substack{a, y \\ a \neq y}} q(a, y|b) = \sum_{\substack{a, y \\ a \neq y}} p(a) q(y|b) = \sum_y \frac{n - 1}{n} q(y|b) = 1 - \varepsilon \,.$$

Therefore, using that $q(x|a,b) = q(x|a)$ and hence $q(x|b{=}0) = q(x|b{=}1)$, we calculate

$$\text{Prob}_q^*[\mathsf{Acc}|0] + \text{Prob}_q^*[\mathsf{Acc}|1]$$
$$= q(x{=}0|b{=}0) + q(x{=}\perp, y \neq a|b{=}0) + q(x{=}1|b{=}1) + q(x{=}\perp, y \neq a|b{=}1)$$
$$\leq q(x{=}0|b{=}0) + q(x{=}1|b{=}0) + q(x{=}\perp|b{=}0) + q(y{\neq}a|b{=}1)$$
$$= 1 + (1 - \varepsilon).$$

We now adapt this argument to $\varepsilon = m/n$, where m and n are integers such that $0 < m \leq n$. For every $a \in [n]$, we define a subset S_a of $[n]$ as

$$S_a = \{a + i \bmod n \mid i \in \{0, \ldots, m-1\}\}.$$

We adapt our scheme by replacing the condition $r = a$ with $r \in S_a$. Clearly, the scheme is still sound. Since every S_a has exactly m elements, the scheme is ε-hiding: the probability that the first prover reveals b is $m/n = \varepsilon$; otherwise, he does not give any information about b. The proof that the scheme is $(1 - \varepsilon)$-binding goes through as before if we can show that $q(y \notin S_a|a,b) = 1 - \varepsilon$ for any non-signaling strategy q. Indeed, for every $y \in [n]$, there are exactly m values for a such that $y \in S_a$. Since $a \in [n]$ is selected randomly and $q(y|a,b)$ is independent of a, we have $q(y \notin S_a|a,b) = 1 - m/n = 1 - \varepsilon$. $\square$

4.2 Arbitrary Schemes

We now remove the restriction on the scheme to be simple. As before, we first consider the case of a perfectly hiding scheme.

Theorem 4.3. *Let* Com *be a single-round two-prover commitment scheme. If* Com *is perfectly hiding, then there exists a non-signaling two-prover strategy* $q(x, x', y, y'|a, a', b, b')$ *such that*

$$\text{Prob}_q^*[\mathsf{Acc}|b] = \text{Prob}_p[\mathsf{Acc}|b]$$

for $b \in \{0, 1\}$.

Proof. Com being perfectly hiding means that $d(p(x_0, x_0'|a, a'), p(x_1, x_1'|a, a')) = 0$ for all a and a'. Gluing together the distributions $p(x_0, x_0', y_0, y_0'|a, a')$ and $p(x_1, x_1', y_1, y_1'|a, a')$ along (x_0, x_0') and (x_1, x_1') for every (a, a'), we obtain a distribution $p(x_0, x_0', x_1, x_1', y_0, y_0', y_1, y_1'|a, a')$ with the correct marginals and $p((x_0, x_0') \neq (x_1, x_1')|a, a') = 0$. That is, we have $x_0 = x_1$ and $x_0' = x_1'$ with certainty. We now define a strategy for dishonest provers as (Figure 4)

$$q(x, x', y, y'|a, a', b, b') := p(x_0, x_0', y_b, y_{b'}'|a, a').$$

Since $p(x_0, x_0', y_b, y_b'|a, a') = p(x_b, x_b', y_b, y_b'|a, a')$, it holds that $\text{Prob}_q^*[\mathsf{Acc}|b] = \text{Prob}_p[\mathsf{Acc}|b]$. It remains to show that this distribution satisfies the non-signaling and causality constraints (C1) up to (NS2) of Definition 3.2. This is done below.

- For (C1), note that summing up over y and y' yields $q(x, x'|a, a', b, b') = p(x_0, x_0'|a, a')$, which indeed does not depend on b and b'.

- For (NS1), note that $q(x, y|a, a', b, b') = p(x_0, y_b|a, a') = p(x_b, y_b|a, a') = p(x_b, y_b|a)$, where the last equality holds by the non-signaling property of $p(x_b, y_b|a, a')$.

- For (C2), first note that
$$q(x, x', y|a, a', b, b') = p(x_0, x_0', y_b|a, a') \tag{2}$$
which does not depend on b'. We then see that (C2) holds by dividing by $q(x, y|a, a', b, b') = p(x_0, y_b|a, a')$.

- For (NS2), divide Equation (2) by $q(x, x'|a, a', b, b') = p(x_0, x_0'|a, a')$

The properties (C1) to (NS2) with the roles of the primed and unprimed variables exchanged follows from symmetry. This concludes the proof. $\square$
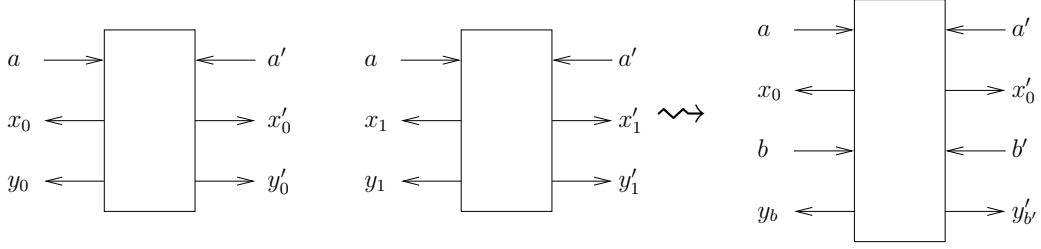
Figure 4: Defining the strategy q from $p(x_0, x'_0, y_0, y'_0 | a, a')$ and $p(x_1, x'_1, y_1, y'_1 | a, a')$ glued together.

The case of non-perfectly hiding schemes is more involved. At first glance, one might expect that by proceeding analogously to the proof of Theorem 4.3 — i.e., gluing together $p(x_0, x'_0, y_0, y'_0 | a, a')$ and $p(x_1, x'_1, y_1, y'_1 | a, a')$ along (x_0, x'_0) and (x_1, x'_1) and defining q the same way — one can obtain a strategy q that succeeds with probability $1 - \varepsilon$ if the scheme is ε-hiding. Unfortunately, this approach fails because in order to show (NS1) we use that $p(x_0, y_1 | a, a') = p(x_1, y_1 | a, a')$ which in general does not hold for commitment schemes that are not perfectly hiding. As a consequence, our proof is more involved, and we have a constant-factor loss in the parameter.

Theorem 4.4. *Let* Com *be a single-round two-prover commitment scheme and suppose that it is ε-hiding. Then there exists a non-signaling two-prover strategy $q(x, x', y, y' | a, a', b, b')$ such that*

$$\mathrm{Prob}^*_q[\mathsf{Acc}|0] = \mathrm{Prob}_p[\mathsf{Acc}|0] \quad and \quad \mathrm{Prob}^*_q[\mathsf{Acc}|1] \geq \mathrm{Prob}_p[\mathsf{Acc}|1] - 5\varepsilon \,.$$

Thus, if Com *is perfectly sound, it is at best $(1 - 5\varepsilon)$-binding.*

To prove this result, we use two lemmas. In the first one, we add the additional assumptions that $p(x_0 | a, a') = p(x_1 | a, a')$ and $p(x'_0 | a, a') = p(x'_1 | a, a')$. The second one shows that we can tweak an arbitrary scheme in such a way that these additional conditions hold. The proofs are given in Appendix B.

Lemma 4.5. *Let* Com *be a ε-hiding two-prover commitment scheme with the additional property that $p(x_0 | a, a') = p(x_1 | a, a')$ and $p(x'_0 | a, a') = p(x'_1 | a, a')$. Then, there is a non-signaling $p'(x_1, x'_1, y_1, y'_1 | a, a')$ such that*

$$d\big(p'(x_1, x'_1, y_1, y'_1 | a, a'), p(x_1, x'_1, y_1, y'_1 | a, a')\big) \leq \varepsilon$$

and $p'(x_1, x'_1 | a, a') = p(x_0, x'_0 | a, a')$.

As usual, the non-signaling requirement on $p'(x_1, x'_1, y_1, y'_1 | a, a')$ is to be understood as $p'(x_1, y_1 | a, a') = p'(x_1, y_1 | a)$ and $p'(x'_1, y'_1 | a, a') = p'(x'_1, y'_1 | a')$.

Lemma 4.6. *Let* Com *be a ε-hiding two-prover commitment scheme. Then, there exists a non-signaling $\tilde{p}(x_1, x'_1, y_1, y'_1 | a, a')$ such that*

$$d\big(\tilde{p}(x_1, x'_1, y_1, y'_1 | a, a'), p(x_1, x'_1, y_1, y'_1 | a, a')\big) \leq 2\varepsilon$$

which has the property that $\tilde{p}(x_1 | a, a') = p(x_0 | a, a')$ and $\tilde{p}(x'_1 | a, a') = p(x'_0 | a, a')$.

With these two lemmas, Theorem 4.4 is easy to prove.

Proof of Theorem 4.4. We start with a ε-hiding non-signaling bit-commitment scheme Com. We apply Lemma 4.6 and obtain a distribution $\tilde{p}(x_1, x'_1, y_1, y'_1 | a, a')$ which is 2ε-close to $p(x_1, x'_1, y_1, y'_1 | a, a')$ and satisfies $\tilde{p}(x_1 | a, a') = p(x_0 | a, a')$ and $\tilde{p}(x'_1 | a, a') = p(x'_0 | a, a')$. Furthermore, by triangle inequality

$$d\big(\tilde{p}(x_1, x'_1 | a, a'), p(x_0, x'_0 | a, a')\big) \leq 3\varepsilon \,.$$

Thus, replacing $p(x_1, x'_1, y_1, y_1 | a, a')$ by $\tilde{p}(x_1, x'_1, y_1, y'_1 | a, a')$ gives us a 3ε-hiding two-prover commitment scheme that satisfies the extra assumption in Lemma 4.5. As a result, we obtain a distribution $p'(x_1, x'_1, y_1, y'_1 | a, a')$ that is 3ε-close to $\tilde{p}(x_1, x'_1, y_1, y'_1 | a, a')$, and thus 5ε-close to $p(x_1, x'_1, y_1, y'_1 | a, a')$, with the property that $p'(x_1, x'_1 | a, a') = p(x_0, x'_0 | a, a')$. Therefore, replacing $\tilde{p}(x_1, x'_1, y_1, y'_1 | a, a')$ by $p'(x_1, x'_1, y_1, y'_1 | a, a')$ gives us a *perfectly hiding* two-prover commitment scheme, to which we can apply Theorem 4.3. As a consequence, there exists a non-signaling strategy $q(x, x', y, y' | a, a')$ with $\mathrm{Prob}^*_q[\mathsf{Acc}|0] = \mathrm{Prob}_p[\mathsf{Acc}|0]$ and $\mathrm{Prob}^*_q[\mathsf{Acc}|1] \geq \mathrm{Prob}_p[\mathsf{Acc}|1] - 5\varepsilon$, as claimed. $\square$

Remark 4.7. *If* Com *already satisfies* $p(x_0|a, a') = p(x_1|a, a')$ *and* $p(x'_0|a, a') = p(x'_1|a, a')$, *we can apply Lemma 4.5 right away and thus get a strategy* q *with* $\mathrm{Prob}^*_q[\mathsf{Acc}|0] = \mathrm{Prob}_p[\mathsf{Acc}|0]$ *and* $\mathrm{Prob}^*_q[\mathsf{Acc}|1] \geq \mathrm{Prob}_p[\mathsf{Acc}|1] - \varepsilon$. *Thus, with this additional condition, we still obtain a tight bound as in Theorem 4.1.*

4.3 Multi-Round Schemes

We briefly discuss a limited extension of our impossibility results for single-round schemes to schemes where during the commit phase, there is multi-round interaction between the verifier V and the two provers P and Q. We still assume the opening phase to be one-round; this is without loss of generality in case of *classical* two-prover commitment schemes (where the honest provers are restricted to be classical). In this setting, we have the following impossibility result, which is restricted to perfectly-hiding schemes.

Theorem 4.8. *Let* Com *be a multi-round two-prover commitment scheme. If* Com *is perfectly hiding, then there exists a non-signaling two-prover strategy that completely breaks the binding property, in the sense of Theorem 4.3.*

A formal proof of this statement requires a definition of n-round non-signaling bipartite systems for arbitrary n. Such a definition can be based on the intuition that it must be possible to query the left and right subsystem in any order. With this definition, the proof is a straightforward extension of the proof of Theorem 4.3: the non-signaling strategy is obtained by gluing together $p(\mathbf{x}_0, \mathbf{x}'_0|\mathbf{a}, \mathbf{a}')$ and $p(\mathbf{x}_1, \mathbf{x}'_1|\mathbf{a}, \mathbf{a}')$ along $(\mathbf{x}_0, \mathbf{x}'_0)$ and $(\mathbf{x}_1, \mathbf{x}'_1)$, and setting $q(\mathbf{x}, \mathbf{x}', y, y'|\mathbf{a}, \mathbf{a}', b, b') := p(\mathbf{x}_0, \mathbf{x}'_0, y_b, y'_{b'}|\mathbf{a}, \mathbf{a}')$, where we use bold-face notation for the vectors that collect the messages sent during the multi-round commit phase: $\mathbf{a}$ collects all the messages sent by the verifier to the prover P, etc.

As far as we see, the proof of the non-perfect case, i.e. Theorem 4.4, does not generalize immediately to the multi-round case. As such, proving the impossibility of *non-perfectly-hiding multi-round* two-prover commitment schemes remains an open problem.

5 Possibility of Three-Prover Commitments

It turns out that we can overcome the impossibility results by adding a third prover. We will describe a scheme that is perfectly sound, perfectly hiding and 2^{-n}-binding with communication complexity $O(n)$. We now define what it means for three provers to be non-signaling; since our scheme is similar to a simple scheme, we can simplify this somewhat. We consider distributions $q(x, y, z|a, b, c)$ where a and x are input and output of the first prover P, b and y are input and output of the second prover Q and c and z are input and output of the third prover R.

Definition 5.1. *A conditional distribution* $q(x, y, z|a, b, c)$ *is called a* non-signaling (one-round) tripartite system *if it satisfies*

$$q(x|a, b, c) = q(x|a) , \quad q(y|a, b, c) = q(y|b) , \quad q(z|a, b, c) = q(z|c) ,$$
$$q(x, y|a, b, c) = q(x, y|a, b) , \quad q(x, z|a, b, c) = q(x, z|a, c)$$
$$and \quad q(y, z|a, b, c) = q(y, z|b, c) .$$

In other words, for any way of viewing q as a bipartite system by dividing in- and outputs consistently into two groups, we get a non-signaling bipartite system. Actually, by means of Lemma A.2, it is not hard to see that the first three requirements follow by the (union of the) latter three.

We restrict to *simple* schemes, where during the commit phase, only P is active, sending x upon receiving a from the verifier, and during the opening phase, only Q and R are active, sending y and z to the verifier, respectively.

Definition 5.2. *A* simple three-prover commitment scheme Com *consists of a probability distribution* $p(a)$, *two distributions* $p_0(x, y, z|a)$ *and* $p_1(x, y, z|a)$, *and an acceptance predicate* $\mathsf{Acc}(x, y, z|a, b)$.
It is called classical/quantum/non-signaling if $p_b(x, y, z|a)$ *is, when understood as a tripartite system* $p_b(x, y, z|a, \emptyset, \emptyset)$ *with two "empty" inputs.*

Soundness and the hiding-property are defined in the obvious way. As for the binding property, for a simple three-prover commitment scheme Com and a non-signaling strategy $q(x, y, z|a, b, c)$, let

$$\mathrm{Prob}^*_q[\mathsf{Acc}|b] = \sum_{a, x, y, z} p(a) \cdot q(x, y, z|a, b, b) \cdot \mathsf{Acc}(x, y, z|a, b) .$$

We say that Com is δ-binding if

$$\mathrm{Prob}_q^*[\mathsf{Acc}|0] + \mathrm{Prob}_q^*[\mathsf{Acc}|1] \leq 1 + \delta.$$

Theorem 5.3. *For every positive integer n, there exists a classical simple three-prover commitment scheme that is perfectly sound, perfectly hiding and 2^{-n}-binding. The verifier communicates n bits to the first prover and receives n bits from each prover.*

The scheme that achieves this is essentially the same as the example two-prover scheme described in the introduction, except that we add a third prover that imitates the actions of the second. To be more precise: the provers P, Q and R have as shared randomness a uniformly random $r \in \{0,1\}^n$. The verifier V chooses a uniformly random $a \in \{0,1\}^n$ and sends it to P. As commitment, P returns $x := r \oplus a \cdot b$. To open the commitment to b, Q and R send $y := r$ and $z := r$ to V who accepts if and only if $y = z$ and $x = y \oplus a \cdot b$.

Before beginning with the formal proof that this scheme has the properties stated in our theorem, we give some intuition. Let a and x be the input and output of the dishonest first prover, P. To succeed, the second prover Q has to produce output $x \oplus a \cdot b$ where b is the second prover's input and the third prover R has to produce $x \oplus a \cdot c$ where c is the third prover's input. Our theorem implies that a strategy which always produces these outputs must be signaling. Why is that the case?

In the game that defines the binding-property, we always have $b = c$, but the dishonest provers must obey the non-signaling constraint even in the "impossible" case that $b \neq c$. Let us consider the XOR of Q's output and R's output in the case that $b \neq c$: we get $(x \oplus a \cdot b) \oplus (x \oplus a \cdot c) = a \cdot b \oplus a \cdot c = a$. But in the non-signaling setting, the joint distribution of Q's and R's output may not depend on a. Thus, the strategy we suggested does not satisfy the non-signaling constraint. Let us now prove the theorem.

Proof of Theorem 5.3. It is easy to see that the scheme is sound. Furthermore, for every fixed a and b, $p_b(x|a)$ is uniform, so the scheme is perfectly hiding. Now consider a non-signaling strategy q for dishonest provers. The provers succeed if and only if $y = z = x \oplus a \cdot b$. Define $q(a, x, y, z|b, c) = p(a) \cdot q(x, y, z|a, b, c)$. The non-signaling property implies that

$$q(y = x \oplus a \cdot b|a, b, c = 0) = q(y = x \oplus a \cdot b|a, b, c = 1) \quad \text{and} \tag{3}$$

$$q(z = x \oplus a \cdot c|a, b = 0, c) = q(z = x \oplus a \cdot c|a, b = 1, c). \tag{4}$$

It follows that

$$\begin{aligned}
&\mathrm{Prob}_q^*[\mathsf{Acc}|0] + \mathrm{Prob}_q^*[\mathsf{Acc}|1] \\
&\quad = q(y = x \oplus a \cdot b, z = x \oplus a \cdot c|b = 0, c = 0) \\
&\qquad + q(y = x \oplus a \cdot b, z = x \oplus a \cdot c|b = 1, c = 1) \\
&\quad \leq q(y = x \oplus a \cdot b|b = 0, c = 0) + q(z = x \oplus a \cdot c|b = 1, c = 1) \\
&\quad = q(y = x \oplus a \cdot b|b = 0, c = 1) + q(z = x \oplus a \cdot c|b = 0, c = 1) \\
&\qquad \text{by Equations (3) and (4)} \\
&\quad \leq 1 + q(y = x \oplus a \cdot b, z = x \oplus a \cdot c|b = 0, c = 1) \text{ by Equation (1)}
\end{aligned}$$

It now remains to upper-bound $q(y = x \oplus a \cdot b, z = x \oplus a \cdot c|b = 0, c = 1)$. Since $p(a)$ is uniform and $q(y, z|a, b, c)$ is independent of a, we have

$$q(y = x \oplus a \cdot b, z = x \oplus a \cdot c|b = 0, c = 1) \leq q(y \oplus z = a|b = 0, c = 1) = \frac{1}{2^n}$$

and thus our scheme is 2^{-n}-binding. $\qquad\square$

Remark 5.4. *The three-prover scheme above has the drawback that two provers are involved in the opening phase; as such, there needs to be agreement on whether to open the commitment or not; if there is disagreement then this may be problematic in certain applications. However, P and Q are not allowed to communicate. One possible solution is to have V forward an authenticated "open" or "not open" message from P to Q and R. This allows for some communication from P to Q and R, but if the size of the authentication tag is small enough compared to the security parameter of the scheme, i.e., n, then security is still ensured.*

Acknowledgements

References

[Bel64]   John Stewart Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[BGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-Prover Interactive Proofs: How to Remove Intractability Assumptions. In Janos Simon, editor, *STOC*, pages 113–131. ACM, 1988.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[CSST11]   Claude Crépeau, Louis Salvail, Jean-Raymond Simard, and Alain Tapp. Two Provers in Isolation. In Dong Hoon Lee and Xiaoyun Wang, editors, *ASIACRYPT*, volume 7073 of *Lecture Notes in Computer Science*, pages 407–430. Springer, 2011.

[DMS00]   Paul Dumais, Dominic Mayers, and Louis Salvail. Perfectly concealing quantum bit commitment from any quantum one-way permutation. In Bart Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000, International Conference on the Theory and Application of Cryptographic Techniques, Bruges, Belgium, May 14-18, 2000, Proceeding*, volume 1807 of *Lecture Notes in Computer Science*, pages 300–315. Springer, 2000.

[Ken99]   Adrian Kent. Unconditionally secure bit commitment. *Phys. Rev. Lett.*, 83:1447–1450, 1999.

[KRR14]   Yael Tauman Kalai, Ran Raz, and Ron D. Rothblum. How to delegate computations: the power of no-signaling proofs. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 485–494. ACM, 2014.

[LC97]   Hoi-Kwong Lo and H. F. Chau. Is quantum bit commitment really possible? *Phys. Rev. Lett.*, 78:3410–3413, Apr 1997.

[May97]   Dominic Mayers. Unconditionally Secure Quantum Bit Commitment is Impossible. *Phys. Rev. Lett.*, 18:3414–3417, 1997.

[PR94]   Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[Ras85]   Peter Rastall. Locality, bell's theorem, and quantum mechanics. *Foundations of Physics*, 15(9):963–972, 1985.

[RK05]   Renato Renner and Robert König. Universally Composable Privacy Amplification Against Quantum Adversaries. In Joe Kilian, editor, *TCC*, volume 3378 of *Lecture Notes in Computer Science*, pages 407–425. Springer, 2005.

[TK85]   Boris S. Tsirelson and Leonid A. Khalfin. Quantum and quasi-classical analogs of Bell inequalities. In *Symposium on the Foundations of Modern Physics*, pages 441–460, 1985.

[Yan13]   Nan Yang. Zero-Knowledge Multi-Prover Interactive Proofs. Master's thesis, Concordia University Montreal, 2013.

A   Capturing the Non-signaling property

In this section, we argue that Definition 3.2 is not only necessary but also sufficient to capture the non-signaling constraint. Consider a two-round bipartite system that conforms to Definition 3.2. We show that the two subsystems can be queried *in any order* without altering the output distribution, as long as the order of rounds for each subsystem individually is respected. Thus, it is impossible to obtain information about the right side of the system by observing only the behaviour on the left side (and vice versa), which shows that Definition 3.2 is indeed sufficient. First, we point out the following.

Remark A.1. *(C1) and (NS1) together imply that $q(x|a, b)$ and $q(x|a, a')$ are well-defined and satisfy*

$$q(x|a, b) = q(x|a) \quad (C3) \qquad and \qquad q(x|a, a') = q(x|a) \qquad (NS3).$$

This follows from Lemma A.2 below.

Lemma A.2. *Any conditional distribution $q(x|a, b, c, d)$ that satisfies $q(x|a, b, c, d) = q(x|a, b)$ as well as $q(x|a, b, c, d) = q(x|a, c)$, must also satisfy $q(x|a, b, c, d) = q(x|a)$.*

Proof. Recall that, by convention, $q(x|a, b, c, d) = q(x|a, b)$ means $q(x|a, b, c, d) = q(x|a, b, c', d')$ for all x, a, b, c, c', d, d', and similarly for $q(x|a, b, c, d) = q(x|a, c)$. As such, for arbitrary $x, a, b, b', c, c', d, d'$ it holds that

$$q(x|a, b, c, d) = q(x|a, b, c', d') = q(x|a, b', c', d')$$

and thus $q(x|a, b, c, d) = q(x|a)$. □

If $q(x, x', y, y'|a, a', b, b')$ is a non-signaling two-round bipartite system, then it can be written as

$$
\begin{aligned}
q(x, x', y, y'|a, a', b, b') &= q(x, y|a, b) \cdot q(x', y'|x, y, a, a', b, b') \\
&= q(x|a) \cdot q(y|x, a, b) \cdot q(x'|x, y, a, a', b) \cdot q(y'|x, y, a, a', b, b')
\end{aligned}
$$

where the first equality uses (NS1), and the second uses (C3) and (C2), and as

$$
\begin{aligned}
&q(x, x', y, y'|a, a', b, b') \\
=\ & q(x, x'|a, a') \cdot q(y, y'|x, x', a, a', b, b') \\
=\ & q(x|a) \cdot q(x'|x, a, a') \cdot q(y|x, x', a, a', b) \cdot q(y'|x, x', y, a, a', b, b')
\end{aligned}
$$

where the first equality uses (C1), and the second uses (NS3) and (NS2), and the second equality can also be replaced by

$$= q(x|a) \cdot q(x'|x, a, a') \cdot q(y'|x, x', a, a', b') \cdot q(y|x, x', y, a, a', b, b').$$

And, similarly, with the roles of the primed and unprimed variables exchanged. This shows that the two subsystems can be queried in any order. For instance, one can first query the left subsystem to get x on input a, distributed according to $q(x|a)$, and then y on input b, distributed according to $q(y|x, a, b)$, and then then one can query the right subsystem twice to get x' and y', distributed according to $q(x'|x, y, a, a', b)$ and $q(y'|x, y, a, a', b, b')$, respectively.[4] Or, one can first query the left subsystem once to obtain x, then query the right subsystem to obtain x' etc. It is straightforward to verify that all six eligible orderings are possible.

B Proofs of Lemma 4.5 and Lemma 4.6

Proof of Lemma 4.5. For arbitrary a and a', we use Lemma 2.4 to glue together the distributions $p(x_0, x'_0, y_0, y'_0|a, a')$ and $p(x_1, x'_1, y_1, y'_1|a, a')$ to obtain a joint distribution $p(x_0, x'_0, x_1, x'_1, y_0, y'_0, y_1, y'_1|a, a')$ such that

$$p\big((x_0, x'_0) \neq (x_1, x'_1)|a, a'\big) \leq \varepsilon,$$

and thus $d(p(x_0, x'_0, y_1, y'_1|a, a'), p(x_1, x'_1, y_1, y'_1|a, a')) \leq \varepsilon$. Let Λ be the event that both $x_0 = x_1$ and $x'_0 = x'_1$. We define $p'(x_1, x'_1, y_1, y'_1|a, a')$ as follows, where x_0 is associated with x_1 and x'_0 with x'_1:

$$
\begin{aligned}
p'(x_1, x'_1, y_1, y'_1|a, a') &:= p(\Lambda, x_0, x'_0|a, a') \cdot p(y_1, y'_1|\Lambda, x_1, x'_1, a, a') \\
&\quad + p(\overline{\Lambda}, x_0, x'_0|a, a') \cdot r(y_1|x_0, a, a') \cdot r(y'_1|x'_0, a, a') \\
&= p(\Lambda, x_1, x'_1, y_1, y'_1|a, a') \\
&\quad + p(\overline{\Lambda}, x_0, x'_0|a, a') \cdot r(y_1|x_0, a, a') \cdot r(y'_1|x'_0, a, a')
\end{aligned}
$$

---

[4]Note that in oder to sample, say, x' according to $q(x'|x, y, a, a', b)$, it seems like that the right subsystem needs to know a, x etc., i.e., that communication is necessary, contradicting the non-signaling requirement. However, this reasoning merely shows that in general, such a non-signaling system is not classical.

where $r(y_1|x_0, a, a')$ and $r(y_1'|x_0', a, a')$ are to be defined later, and the last equality holds by definition of Λ.[5]

The claim about the closeness to $p(x_1, x_1', y_1, y_1'|a, a')$ follows from the fact that $p(\overline{\Lambda}|a, a') \leq \varepsilon$. Furthermore, we have $p'(x_1, x_1'|a, a') = p(\Lambda, x_0, x_0'|a, a') + p(\overline{\Lambda}, x_0, x_0'|a, a') = p(x_0, x_0'|a, a')$ as claimed.

It remains to show that we can achieve p' to be non-signaling. For that, we simply define $r(y_1|x_0, a, a')$, and similarly $r(y_1'|x_0', a, a')$, in such a way that $p'(x_1, y_1|a, a') = p(x_1, y_1|a, a')$; this does the job since $p(x_1, y_1|a, a') = p(x_1, y_1|a)$, and as such $p'(x_1, y_1|a, a') = p'(x_1, y_1|a)$. Note that

$$p'(x_1, y_1|a, a') = p(\Lambda, x_1, y_1|a, a') + p(\overline{\Lambda}, x_0|a, a') \cdot r(y_1|x_0, a, a'). \tag{5}$$

Thus, we set

$$r(y_1|x_0, a, a') := \frac{p(x_1, y_1|a, a') - p(\Lambda, x_1, y_1|a, a')}{p(\overline{\Lambda}, x_0|a, a')} = \frac{p(\overline{\Lambda}, x_1, y_1|a, a')}{p(\overline{\Lambda}, x_0|a, a')}$$

It remains to show that $r(y_1|x_0, a, a')$ as defined is indeed a probability distribution, and that things work out also in case $p(\overline{\Lambda}, x_0|a, a') = 0$.

In the latter case, we have $p'(x_1, y_1|a, a') = p(\Lambda, x_1, y_1|a, a')$, independent of the choice of r; thus, it remains to show that $p(\Lambda, x_1, y_1|a, a') = p(x_1, y_1|a, a')$. For that, we observe that $p(\Lambda, x_1|a, a') = p(\Lambda, x_0|a, a') = p(x_0|a, a') = p(x_1|a, a')$, where the first equality is due to the definition of Λ and the last holds by our additional assumption on Com. It follows that

$$\sum_{y_1} p(\Lambda, x_1, y_1|a, a') = p(\Lambda, x_1|a, a') = p(x_1|a, a') = \sum_{y_1} p(x_1, y_1|a, a')$$

and since $p(\Lambda, x_1, y_1|a, a') \leq p(x_1, y_1|a, a')$, it holds that $p(\Lambda, x_1, y_1|a, a') = p(x_1, y_1|a, a')$ as required.

Finally, to show that $r(y_1|x_0, a, a')$ is a probability distribution, we observe that $r(y_1|x_0, a, a') \geq 0$, and, summing over y_1 and using that $p(x_0|a, a') = p(x_1|a, a')$, we see that

$$\sum_{y_1} r(y_1|x_0, a, a') = \frac{p(x_1|a, a') - p(\Lambda, x_1|a, a')}{p(\overline{\Lambda}, x_0|a, a')} = \frac{p(x_0|a, a') - p(\Lambda, x_0|a, a')}{p(\overline{\Lambda}, x_0|a, a')}$$

$$= \frac{p(\overline{\Lambda}, x_0|a, a')}{p(\overline{\Lambda}, x_0|a, a')}$$

$$= 1.$$

In the same way, it is possible to choose $r(y_1'|x_0', a, a')$ so that $p'(x_1', y_1'|a, a') = p(x_1', y_1'|a, a') = p(x_1', y_1'|a')$, using the assumption that $p(x_0'|a, a') = p(x_1'|a, a')$. This concludes the proof. $\qquad\square$

Proof of Lemma 4.6. We begin by adjusting the distribution of x_1. By the hiding property of Com, $p(x_0, x_0'|a, a')$ and $p(x_1, x_1'|a, a')$ are ε-close, and thus in particular $d(p(x_0|a, a'), p(x_1|a, a')) \leq \varepsilon$. Gluing together the distributions $p(x_0|a, a')$ and $p(x_1, x_1', y_1, y_1'|a, a')$ along x_0 and x_1, we get $p(x_0, x_1, x_1', y_1, y_1'|a, a')$ such that

$$p'(x_1, x_1', y_1, y_1'|a, a') := p(x_0, x_1', y_1, y_1'|a, a')$$

satisfies $d(p'(x_1, x_1', y_1, y_1'|a, a'), p(x_1, x_1', y_1, y_1'|a, a')) \leq \varepsilon$ and also $p'(x_1|a, a') = p(x_0|a, a')$.

We show that p' is non-signaling. Since $p'(x_1', y_1'|a, a') = p(x_1', y_1'|a, a')$ and p is non-signaling, it follows that $p'(x_1', y_1'|a, a') = p'(x_1', y_1'|a')$. Showing that $p'(x_1, y_1|a, a') = p'(x_1, y_1|a)$ is equivalent to showing that $p(x_0, y_1|a, a') = p(x_0, y_1|a)$. By the observation in Remark 2.7, the marginal $p(x_0, x_1, y_1|a, a')$ is obtained by gluing together $p(x_0|a, a')$ and $p(x_1, y_1|a, a')$ along x_0 and x_1. Since Com is non-signaling, it holds that $p(x_0|a, a') = p(x_0|a)$ and $p(x_1, y_1|a, a') = p(x_1, y_1|a)$. It follows that $p(x_0, x_1, y_1|a, a') = p(x_0, x_1, y_1|a)$, and therefore that $p(x_0, y_1|a, a') = p(x_0, y_1|a)$.

In order to obtain $\tilde{p}$ as claimed, we repeat the above process. Note that the modification from p to p' did not change the distribution of x_1', y_1', i.e., $p'(x_1', y_1'|a, a') = p(x_1', y_1'|a, a')$, and thus in particular $d(p(x_0'|a, a'), p'(x_1'|a, a')) = d(p(x_0'|a, a'), p(x_1'|a, a')) \leq \varepsilon$. Therefore, exactly as above, we can now adjust the distribution of x_1' in p' and obtain a non-signaling $\tilde{p}(x_1, x_1', y_1, y_1'|a, a')$ that is ε-close to $p'(x_1, x_1', y_1, y_1'|a, a')$ and thus 2ε-close to $p(x_1, x_1', y_1, y_1'|a, a')$, and which satisfies $\tilde{p}(x_1'|a, a') = p(x_0'|a, a')$ and $\tilde{p}(x_1|a, a') = p'(x_1|a, a') = p(x_0|a, a')$, as claimed. $\qquad\square$

---

[5]Algorithmically, the distribution p' should be understood as follows. First, x_0, x_0', x_1 and x_1' are sampled according to the glued-together distribution p. Then, if the event Λ occurred (i.e. $x_0 = x_1$ and $x_0' = x_1'$), y_1 and y_1' are sampled according to the corresponding conditional distribution; otherwise, they are chosen *independently* according to distributions that depend only on x_0 and x_0', respectively.