Cryptographically Blinded Games: Leveraging Players' Limitations for Equilibria and Profit

Pavel Hubáček[1] and Sunoo Park[2]

[1]Aarhus University
[2]MIT

November 13th, 2014

Abstract

In this work we apply methods from cryptography to enable any number of mutually distrusting players to implement broad classes of mediated equilibria of strategic games without the need for trusted mediation.

Our implementation makes use of a (standard) pre-play "cheap talk" phase, in which players engage in free and non-binding communication prior to playing in the original game. In our cheap talk phase, the players execute a secure multi-party computation protocol to sample an action profile from an equilibrium of a "cryptographically blinded" version of the original game, in which actions are encrypted. The essence of our approach is to exploit the power of encryption to selectively restrict the information available to players about sampled action profiles, such that these desirable equilibria can be stably achieved. In contrast to previous applications of cryptography to game theory, this work is the first to employ the paradigm of using encryption to allow players to benefit from hiding information *from themselves*, rather than from others; and we stress that rational players would *choose* to hide the information from themselves.

Keywords. Cheap talk, encryption, mediated equilibria, multi-party computation.

1 Introduction

Nash equilibrium [Nas50] and correlated equilibrium [Aum74] are important solution concepts that have been extensively studied in both traditional and computational game-theoretic contexts. Coarse correlated equilibrium [MV78] is a closely related concept that was proposed as a generalization of correlated equilibrium, which can be more powerful in some settings such as potential games.

In this work we construct protocols for mutually distrusting players to implement any coarse correlated equilibrium (and therefore any correlated equilibrium) of a strategic game without trusted mediation, via cryptographic cheap talk protocols. Our approach draws upon cryptography in two ways: first, we introduce an intermediate, "cryptographically blinded" game from which the players sample according to the desired equilibrium; and second, this sampling is achieved using a secure multi-party computation protocol. Our results address both the computational and perfect (information-theoretic) settings.

This is a working paper, of which an extended abstract appeared in the 15th ACM Conference on Economics and Computation (EC 2014).

Correlated equilibrium. Suppose a mediator samples an action profile a from a known distribution α, and gives as "advice" to each player i his action a_i in a. The distribution α is a correlated equilibrium if, *having seen his advice*, and believing that all other players will follow their advice, no player has incentive to unilaterally deviate from the advice profile. [Aum74] showed that correlated equilibria can achieve higher expected payoffs than Nash equilibria.

Coarse correlated equilibrium. Coarse correlated equilibria are a generalization of correlated equilibria which invokes a notion of commitment. In the mediated scenario described above, α is a coarse correlated equilibrium if no player has incentive not to "promise" or "commit" in advance – *before seeing his advice* a_i – to play according to the advice, as long as he believes that all other players will commit to do the same. Note that if a player does not commit, then he will not see the advice at all, and must therefore play an independent strategy: this is in contrast to correlated equilibria, where deviations may depend on the received advice.

[MRG13] showed that there is a class of potential games in which the Nash equilibrium payoffs can be improved upon by coarse correlated equilibria but not by correlated equilibria (e.g. the Cournot duopoly and public good provision games).

Example 1.1. Let us give a brief example to illustrate the gap between the two types of equilibria. Suppose Alice plays a game Γ where she has a "safe strategy" for which her payoff is always zero. Let α be a distribution over action profiles of Γ, and suppose Alice's expected payoff from α is very high, say, a million dollars – however, some action profiles from α will give her negative payoff. Now, when Alice receives her advice from the mediator, she might be able to deduce that her payoff in the advised action profile will be negative. If this is the case, she will choose to deviate to her safe strategy, so α is not a correlated equilibrium. However, α may still be a coarse correlated equilibrium if Alice can commit before seeing her advice; and importantly, α may be very desirable from Alice's (risk-neutral) point of view, since expected payoff is high.

1.1 Our results

In this work we address the following question:

How can the players of a strategic game implement any coarse correlated equilibrium via (cryptographic) pre-play communication without trusting each other or a mediator?

In the computational setting, we give an implementation for general strategic games, in the form of an extended game comprising a *cryptographic protocol* in the pre-play phase, which securely samples an action profile for a "cryptographically blinded" version of the original game, followed by play in the original game. The blinded game's action space consists of *encryptions* of the original game's actions.

Our implementation has the strong property that any computational coarse correlated equilibrium of the original game corresponds to a payoff-equivalent Nash equilibrium of the extended game. Furthermore, it achieves *strategic equivalence* to the original game, in that every computational Nash equilibrium of the extended game corresponds to a computational coarse correlated equilibrium of the original game. Pre-play communication is via broadcast, as is standard in the cheap talk literature.

In the information-theoretic setting, we give an implementation for strategic games with four or more players, using a similar format of a cryptographically blinded pre-play phase followed by (simultaneous) play in the original game, given private pairwise communication channels between players. As in the computational setting, we achieve strategic equivalence. Both the

restriction to four or more players and the need for a stronger communication model than broadcast are unavoidable, as shown by impossibility results of [Bár92; AH03] which will be discussed in more detail in the next section.

None of our constructions require trusted mediation. After the pre-play phase is complete, there is a single step in which the players invoke a *verifiable proxy* to play the original game according to their instructions. Verifiable parties were introduced in [ILM11], and will be detailed further in Section 1.2. No trust need be placed in the verifiable proxy, because anyone can check whether ir has acted correctly; and we stress that unlike the usual mediator for coarse correlated equilibria, the verifiable proxy does not communicate anything to the players which may *affect their strategies* in the game. Informally, it simply performs a "translation" of a player's chosen strategy from one form into another.

Finally, our constructions require *no physical assumptions* and can be executed entirely over a distributed network. This contrasts with a number of previous works such as [LMPS04; ILM11] which require "physical envelopes".

1.2 Relation to prior work

Cheap talk. The *pre-play* literature considers the general problem of implementing equilibria without mediation, as follows: given an abstract game Γ, the aim is to devise a concrete communication game Γ' having an equilibrium that is payoff-equivalent to a desirable equilibrium in Γ, where the concrete game may have a pre-play *cheap talk* phase in which players engage in communication that is neither costly nor binding, and has no impact on players' payoffs except insofar as it may influence future actions. In the literature there has been much focus on implementing correlated equilibria [Bár92; BP98; AH03].

Power of commitment. It has long been recognized that the possibility to *commit* to strategies in advance can increase the payoffs achievable in a game, starting with the work of [vS34], who proposed a leader/follower structure to games where the leader moves first (and thereby "commits" to his strategy). [SZ10] showed that transforming a strategic game into a leader/follower form allows the leader (i.e. the committer) to do at least as well as in the Nash and correlated equilibria of the strategic game. Moreover, they show that coarse correlated equilibria, with their arguably stronger notion of commitment, can yield higher payoffs than the leader/follower transformation. More recently, [LKC12] studied the advantage of commitment from a quantitative perspective and showed that the extremal "value of commitment" is in fact unbounded in many classes of games.

In this work, we achieve the payoffs of coarse correlated equilibria without resorting to the assumption of binding contracts: instead, we use the power of encryption to hide information that, if known to the players, could render the situation unstable. We stress that the players are *given the choice*, rather than forced, to hide information from themselves – and we find that it is in their rational interest to do so since coarse correlated equilibria can offer high payoffs.

Cryptographic cheap talk and computational equilibria. [DHR00] introduced the idea of *cryptographic cheap talk*, in which players execute a cryptographic protocol during the pre-play phase; and they defined *computational equilibria*, which are solution concepts stable for computationally bounded (probabilistic polynomial time) players who are indifferent to negligible gains. Their cryptographic cheap talk protocols efficiently implement some computational correlated equilibria of two-player games. Moreover, their notion of computational equilibria suffers from *empty threats* (Definition D.1), which cause instability for sequentially rational players in the pre-play game. This was partially addressed by a new solution concept of [GLR10];

however, [HNR13] subsequently showed that in general, correlated equilibria cannot be achieved without empty threats by (cryptographic) cheap talk.

Our results in the computational setting use the equilibrium definitions of [DHR00]; however, in our "cryptographically blinded" games, empty threats cannot occur. By converting games into blinded games, our constructions implement all coarse correlated equilibria without empty threats: this comes at the cost of a single mediated "translation" step using a third party, discussed in the next paragraph. We consider this step to be a "necessary" and mild requirement given that the impossibility result of [HNR13] renders some additional assumption necessary to achieve all (coarse) correlated equilibria without empty threats.

Removing trusted mediation. Removing the need to trust a mediator in the implementation of equilibria and mechanisms has long been a subject of interest in game theory and cryptography. The notion of *verifiable mediation* was introduced by [ILM11], who highlighted the difference between the usual concept of a *trusted mediator*, and the weaker concept of a *verifiable mediator* who performs actions in a publicly verifiable way and without possessing any information that should be kept secret. Recent applications of verifiable mediation include the strong correlated equilibrium implementation of [ILM11], and the rational secret sharing scheme of [MS09].

In this paper, we introduce the new notion of a *verifiable proxy*. As in verifiable mediation, the actions of a verifiable proxy are publicly verifiable. However, our notion is incomparable to [ILM11]'s verifiable mediation, because:

- a verifiable proxy for a strategic game does not give the players any information that affects their strategic choices in the game; and
- a verifiable proxy may possess information that should be kept secret.

More discussion about the merits of these definitions is given in Section 4.2.

As a simple illustration, consider a sealed-bid auction: much more trust is placed in a mediator who collects all the players' bids and just announces the winner, than in a mediator who collects the bids, opens them publicly, and allows everyone to compute the outcome themselves.

In our setting, the verifiable proxy performs a single "translation" step on behalf of the players, at the end of the pre-play phase, in which it takes strategies submitted by the players and "translates" them into a different format. In particular, the proxy acts independently and identically with respect to each player, and therefore is not implementing the correlation aspect.

Strategic equivalence property. An important concern in implementation theory is the strategic equivalence of an implementation to the underlying game: it is desirable that implementations have the "same" equilibria as the underlying game, and in particular do not introduce new ones. This was first considered by the *full implementation* concept of [Mas99], and extended by subsequent works such as [ILM11] who proposed a stronger notion of *perfect implementation* for certain games. Although this literature is not directly applicable to the present work (as our results lie in the pre-play realm), we extend these ideas and find that the cheap talk extensions of our cryptographically blinded games achieve "best possible" strategic equivalence in that their Nash equilibria correspond exactly to the coarse correlated equilibria of their underlying games. This strategic equivalence notion is "best possible" in the sense that in the pre-play setting, the possibility of arbitrary communication in the pre-play phase inherently introduces the possibility of additional equilibria compared to the simpler one-shot game. Interestingly, Alwen et al. [Alw+09] showed that a very strong notion of strategic equivalence can be achieved if communication in the pre-play is restricted in such a way that players cannot

communicate directly with each other, but only through a mediator who may "censor" some of the communication.

Computationally unbounded setting. To our knowledge, existing work in applying cryptographic tools to game theory has focused overwhelmingly on the setting of computationally bounded players and computational equilibria. In contrast, we consider the computationally unbounded setting too. Our result for the computational setting is stronger and more efficient than our information-theoretic solution: in particular, the computational result holds for games with any number of players, and requires only a broadcast channel for communication between players.

In the computationally unbounded setting it was proven by [Bár92] that correlated equilibria cannot be achieved by cheap talk between fewer than four players, and indeed, this fits neatly with a more general result of [BGW88; CCD88] in the context of secure protocols. Accordingly, our information-theoretic results only apply for games of four or more players; however, improving on the protocols of [Bár92], we achieve not only correlated equilibria but coarse correlated equilibria for all games of this type.

Furthermore, in the computationally unbounded setting it has been proven [AH03] that communication by broadcast alone is *insufficient* to achieve (non-trivial) correlated equilibria by cheap talk, so our result is of interest notwithstanding its stronger requirement of private communication channels between players. Indeed, the private-channels model has been extensively studied in both distributed computing (e.g. [FLP85; KDG03]) and multi-party computation (e.g. [BGW88; CCD88]) as an interesting strengthening of the communication model that allows for much stronger and/or more efficient protocols than the broadcast model. We therefore consider it natural and compelling to apply this model in the game-theoretical setting.

1.3 Organization

In Sections 2 and 3 we provide game-theoretical and cryptographic background. In Section 4 we introduce cryptographically blinded games. These are the essential building block for the cheap talk protocols detailed in Section 5 that implement all coarse correlated equilibria of general strategic games. At the end of Section 5 we discuss the efficiency of our protocols.

1.4 Notation

For $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \ldots, n\}$. For a set S, let $\mathcal{P}(S)$ denote the powerset of S, and let $\Delta(S)$ denote the set of all distributions over S. Let $s \leftarrow S$ denote that s is a random element of S. Let $\sqcup$ denote the disjoint union operation. We write PPT to mean probabilistic polynomial time, and we call distributions that can be sampled in probabilistic polynomial time "PPT-samplable". Let negl denote a negligible function (which tends to zero faster than any inverse polynomial).

2 Game-theoretic background

Definition 2.1 (Finite strategic game)**.** *A finite strategic game $\Gamma = \langle N, (A_i), (u_i) \rangle$ is defined by a finite set N of players, and for each player $i \in N$, a non-empty set of possible actions A_i and a utility function $u_i : \times_{j \in N} A_j \to \mathbb{R}$.*

We refer to an *action profile* $a = (a_j)_{j \in N}$ of a game as an *outcome*, and denote by A the set of outcomes $\times_{j \in N} A_j$. For a given outcome a, we write a_{-i} to denote $(a_j)_{j \in N, j \neq i}$, that is, the

profile of actions of all players other than i; and we use (a'_i, a_{-i}) to denote the action profile where player i's action is a'_i and all other players' actions are as in a.

2.1 Equilibrium concepts

Definition 2.2 (Nash equilibrium)**.** *A* Nash equilibrium *of strategic game* $\Gamma = \langle N, (A_i), (u_i)\rangle$ *is a product distribution* $\alpha \in \times_{j\in N}\Delta(A_j)$ *such that for every player* $i \in N$ *and for all* $a_i^* \in A_i$

$$\mathop{\mathbf{E}}_{a\leftarrow\alpha}[u_i(a)] \geq \mathop{\mathbf{E}}_{a\leftarrow\alpha}[u_i(a_i^*, a_{-i})].$$

Definition 2.3 (Correlated equilibrium)**.** *A* correlated equilibrium *of strategic game* $\Gamma = \langle N, (A_i), (u_i)\rangle$ *is a probability distribution* $\alpha \in \Delta(\times_{j\in N}A_j)$ *such that for every player* $i \in N$*, and for all* $b_i, a_i^* \in A_i$ *satisfying* $\Pr_{a\leftarrow\alpha}[a_i = b_i] > 0$*,*

$$\mathop{\mathbf{E}}_{a\leftarrow\alpha}[u_i(a)|a_i = b_i] \geq \mathop{\mathbf{E}}_{a\leftarrow\alpha}[u_i(a_i^*, a_{-i})|a_i = b_i].$$

Definition 2.4 (Coarse correlated equilibrium)**.** *A* coarse correlated equilibrium *of strategic game* $\Gamma = \langle N, (A_i), (u_i)\rangle$ *is a probability distribution* $\alpha \in \Delta(\times_{j\in N}A_j)$ *such that for every player* $i \in N$ *and for all* $a_i^* \in A_i$

$$\mathop{\mathbf{E}}_{a\leftarrow\alpha}[u_i(a)] \geq \mathop{\mathbf{E}}_{a\leftarrow\alpha}[u_i(a_i^*, a_{-i})].$$

The model of coarse correlated equilibrium allows the players either to "commit in advance" to play according to the mediator's advice (no matter what it turns out to be), or to play an *independent* strategy without learning the advice. A probability distribution is a coarse correlated equilibrium if no player has an incentive to not commit to play according to the mediator's advice.

Because of linearity of expectation, it is sufficient for these equilibrium definitions to consider only deviations to pure strategies. Note that any Nash equilibrium is a correlated equilibrium, and any correlated equilibrium is a coarse correlated equilibrium.

2.2 Computational equilibrium concepts

The following definitions of computational equilibria extend those introduced by [DHR00]. In the computational setting a strategic game induces a family of games parametrized by the security parameter, i.e. $\Gamma = \{\langle N, (A_i^{(k)}), (u_i^{(k)})\rangle\}_{k\in\mathbb{N}}$. Hence, the corresponding solution concepts are ensembles of probability distributions, and the security parameter captures the intuition that players are limited to efficiently computable (PPT) strategies and indifferent to gains negligible in k.

Definition 2.5 (Computational Nash equilibrium)**.** *A* computational Nash equilibrium *of computational strategic game* $\Gamma = \{\langle N, (A_i^{(k)}), (u_i^{(k)})\rangle\}_{k\in\mathbb{N}}$ *is a* PPT*-samplable ensemble of product distributions* $\alpha = \{\alpha^{(k)} = \times_{j\in N}\alpha_j^{(k)}\}_{k\in\mathbb{N}}$ *on* $\{\times_{j\in N}A_j^{(k)}\}_{k\in\mathbb{N}}$ *such that for all players* $i \in N$ *and every* PPT*-samplable ensemble* $\hat{\alpha}_i = \{\hat{\alpha}_i^{(k)}\}_{k\in\mathbb{N}}$ *on* $\{A_i^{(k)}\}_{k\in\mathbb{N}}$*, there exists a negligible* $\varepsilon(\cdot)$ *such that for all large enough* $k \in \mathbb{N}$ *it holds that*

$$\mathop{\mathbf{E}}_{a\leftarrow\alpha^{(k)}}[u_i^{(k)}(a)] \geq \mathop{\mathbf{E}}_{a\leftarrow\alpha^{(k)}, \hat{a}_i\leftarrow\hat{\alpha}_i^{(k)}}[u_i^{(k)}(\hat{a}_i, a_{-i})] - \varepsilon(k).$$

Definition 2.6 (Computational correlated equilibrium)**.** *A* computational correlated equilibrium *of computational strategic game* $\Gamma = \{\langle N, (A_i^{(k)}), (u_i^{(k)})\rangle\}_{k\in\mathbb{N}}$ *is a* PPT*-samplable probability ensemble* $\alpha = \{\alpha^{(k)}\}_{k\in\mathbb{N}}$ *on* $\{\times_{j\in N}A_j^{(k)}\}_{k\in\mathbb{N}}$ *such that for all players* $i \in N$ *and every*

PPT-*samplable ensemble* $\hat{\alpha}_i = \{\hat{\alpha}_i^{(k)}\}_{k\in\mathbb{N}}$ *on* $\{A_i^{(k)}\}_{k\in\mathbb{N}}$ *there exists a negligible* $\varepsilon(\cdot)$ *such that for all large enough* $k \in \mathbb{N}$ *it holds that*

$$\mathop{\mathbf{E}}_{a\leftarrow\alpha^{(k)}}[u_i^{(k)}(a)] \geq \mathop{\mathbf{E}}_{a\leftarrow\alpha^{(k)},\hat{a}_i\leftarrow\hat{\alpha}_i^{(k)}(a_i)}[u_i^{(k)}(\hat{a}_i, a_{-i})] - \varepsilon(k).$$

Definition 2.7 (Computational coarse correlated equilibrium)**.** *A* computational coarse correlated equilibrium *of computational strategic game* $\Gamma = \{\langle N, (A_i^{(k)}), (u_i^{(k)})\rangle\}_{k\in\mathbb{N}}$ *is a* PPT*-samplable probability ensemble* $\alpha = \{\alpha^{(k)}\}_{k\in\mathbb{N}}$ *on* $\{\times_{j\in N} A_j^{(k)}\}_{k\in\mathbb{N}}$ *such that for all players* $i \in N$ *and every* PPT*-samplable ensemble* $\hat{\alpha}_i = \{\hat{\alpha}_i^{(k)}\}_{k\in\mathbb{N}}$ *on* $\{A_i^{(k)}\}_{k\in\mathbb{N}}$*, there exists a negligible* $\varepsilon(\cdot)$ *such that for all large enough* $k \in \mathbb{N}$ *it holds that*

$$\mathop{\mathbf{E}}_{a\leftarrow\alpha^{(k)}}[u_i^{(k)}(a)] \geq \mathop{\mathbf{E}}_{a\leftarrow\alpha^{(k)},\hat{a}_i\leftarrow\hat{\alpha}_i^{(k)}}[u_i^{(k)}(\hat{a}_i, a_{-i})] - \varepsilon(k).$$

Note that in the above definition of computational coarse correlated equilibrium the output of $\hat{\alpha}_i^{(k)}$ is independent of a_i, unlike in the definition of computational correlated equilibrium.

Remark. In later sections we apply the above computational solution concepts in a straightforward way to classical strategic games. For a finite strategic game $\Gamma = \langle N, (A_i), (u_i)\rangle$ we consider the computational version $\{\Gamma^{(k)}\}_{k\in\mathbb{N}}$, where $\Gamma^{(k)} = \Gamma$ for all $k \in \mathbb{N}$. The action space and the utility function do not change with the security parameter in this computational version of Γ; however, the players are limited to efficient (PPT) strategies.

Remark. In the classical setting, it is implicit that the players of a game have oracle access to the utility functions u_i, that is, players can query u_i on any action profile in constant time[1]. Our results apply to all strategic games in the classical setting: hence the requirement that the security parameter be polynomial in the size of the game (i.e. we ensure that players are able to perform the standard task of reading the payoff matrix). With computationally bounded players, however, it seems very natural to consider the case in which computing u_i takes more time. To our knowledge, this difference has been recognized (e.g., [DHR00]) but not much analyzed in the literature; however, it is an important underlying idea of the present work.

2.3 Extensive games

Definitions of extensive form games and subgames are given in Appendix A, along with corresponding equilibrium concepts for the standard and computational settings.

3 Cryptographic background

3.1 Encryption schemes

Our constructions will make use of secret-key and public-key encryption schemes, which are defined below. Note that encryption schemes are parametrized by a security parameter k that determines the "security level" of the scheme.

Definition 3.1 (Secret-key encryption scheme)**.** *A* secret-key encryption scheme *over a message space* $\mathcal{M}$ *is a tuple of* PPT *algorithms* $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ *satisfying the following. Let the ciphertext space be the codomain of* SEnc *and be denoted by* $\mathcal{C}$*.*

[1]Other parameters of the original game, such as the correlated equilibrium distribution, are also assumed to be computable in constant time.

- *The key generation algorithm* SGen *takes no input and outputs a secret key* sk *according to some distribution (inherent to* Σ*). This is denoted by* $sk \leftarrow \mathsf{SGen}()$.

- *The encryption algorithm* SEnc *takes as input a message* $m \in \mathcal{M}$ *and a secret key* sk*, and outputs a ciphertext* $c \in \mathcal{C}$*. This is denoted by* $c \leftarrow \mathsf{SEnc}_{sk}(m)$.

- *The decryption algorithm* SDec *is a deterministic algorithm that takes as input a ciphertext* c *and a secret key* sk*, and outputs a decryption* $m' \in \mathcal{M}$*. This is denoted by* $m' = \mathsf{SDec}_{sk}(c)$.

- *The decryption is always correct, i.e. for every security parameter* k*, and every* $sk \leftarrow \mathsf{SGen}()$ *it holds for every* $m \in \mathcal{M}$ *that* $\mathsf{SDec}_{sk}(\mathsf{SEnc}_{sk}(m)) = m$.

Definition 3.2 (Public-key encryption scheme)**.** *A* public-key encryption scheme *over a message space* $\mathcal{M}$ *is a tuple of* PPT *algorithms* $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ *satisfying the following. Let the ciphertext space be the codomain of* PEnc *and be denoted by* $\mathcal{C}$.

- *The key generation algorithm* PGen *takes input* 1^k*, where* k *is the security parameter, and outputs a public key and secret key pair* (pk, sk).

- *The encryption algorithm* PEnc *takes as input a message* $m \in \mathcal{M}$ *and a public key* pk *and outputs a ciphertext* $c \in \mathcal{C}$.

- *The decryption algorithm* PDec *is a deterministic algorithm that takes as input a ciphertext* c *and a secret key* sk*, and outputs a decryption* $m' \in \mathcal{M}$.

- *The decryption is always correct, i.e. for every security parameter* k*, and every* $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$ *it holds for every* $m \in \mathcal{M}$ *that* $\mathsf{PDec}_{sk}(\mathsf{PEnc}_{pk}(m)) = m$.

3.2 Security definitions

Here we define the following two standard security notions: perfect (information-theoretic) security, and computational security against chosen-ciphertext attacks. The latter is commonly referred to as CCA-security, and is the de facto standard for security of public-key encryption; the former is canonical in the information-theoretic setting.

Remark. Our constructions make use of perfectly secure secret-key encryption and CCA-secure public-key encryption. For convenience, therefore, the security definitions given below refer to secret- and public-key schemes respectively. However, both security definitions may be straightforwardly adapted to apply to both types of encryption (although it is well known that perfect security is impossible in the public-key setting).

Definition 3.3 (Perfectly secure secret-key encryption)**.** *A secret-key encryption scheme* $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ *is* perfectly secure *if for all messages* $m_0, m_1 \in \mathcal{M}$ *and ciphertexts* $c \in \mathcal{C}$*, it holds that* $\Pr[\mathsf{SDec}(\mathsf{SEnc}(m_0)) = m_0] = 1$ *and*

$$\Pr_{sk \leftarrow \mathsf{SGen}()}[\mathsf{SDec}_{sk}(c) = m_0] = \Pr_{sk \leftarrow \mathsf{SGen}()}[\mathsf{SDec}_{sk}(c) = m_1].$$

An alternative and equivalent definition is that a perfectly secure encryption scheme produces ciphertexts that are independent of the messages that they encrypt.

Next, we shall define CCA-security for public-key encryption schemes. The security definition is based on the following experiment, which may be considered to be a game played between a malicious adversary $\mathcal{A}$ and an honest challenger.

The CCA indistinguishability experiment $\mathsf{PubK}^{\mathsf{CCA}}_{\mathcal{A},\Pi}(k)$:

1. The challenger generates a key pair $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$, and sends $(1^k, pk)$ to $\mathcal{A}$.
2. $\mathcal{A}$ has oracle access to PDec_{sk}, and outputs messages $m_0, m_1 \in \mathcal{M}$ of the same length.
3. The challenger samples $b \leftarrow \{0,1\}$, then computes $c \leftarrow \mathsf{PEnc}_{pk}(m_b)$, and sends c to $\mathcal{A}$.
4. $\mathcal{A}$ still has oracle access to PDec_{sk}, but cannot query $\mathsf{PDec}_{sk}(c)$. $\mathcal{A}$ now outputs a bit b'.
5. The output of the experiment is 1 if $b' = b$, and 0 otherwise.

Informally, the adversary "wins the game" if he guesses correctly which of the two messages was encrypted. Clearly, he can win with probability 1/2 by random guessing. The definition of CCA-security formalizes the intuition that he should not be able to do better than that.

Definition 3.4 (CCA-secure public-key encryption)**.** *A public-key encryption scheme* $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ *is* CCA-secure *(i.e. secure against chosen-ciphertext attacks), if for all* PPT *adversaries* $\mathcal{A}$*,* $\Pr[\mathsf{PubK}^{\mathsf{CCA}}_{\mathcal{A},\Pi}(k) = 1] \leq 1/2 + \varepsilon(k)$ *for some negligible* ε*.*

3.3 Non-malleable encryption

Non-malleable encryption was introduced by [DDN00] in the computational setting, and extended to the information-theoretic setting by [HSHI02]. Informally, non-malleability requires that given a ciphertext c, an adversary (who does not know the secret key or the message encrypted by c) cannot generate a different ciphertext c' such that the respective messages are related by some known relation R.

We begin with the simpler information-theoretic definition. Note that [HSHI02] also give a construction of perfectly non-malleable secret-key encryption.

Definition 3.5 (Perfect non-malleability)**.** *A secret-key encryption scheme* $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ *is* perfectly non-malleable *if for all* $c, c', c'' \in \mathcal{C}$ *such that* $c' \neq c \neq c''$ *and all relations* $R : \mathcal{M} \times \mathcal{M} \to \{0,1\}$*,*

$$\Pr_{sk \leftarrow \mathsf{SGen}()}[R(\mathsf{SDec}(c), \mathsf{SDec}(c')) = 1] = \Pr_{sk \leftarrow \mathsf{SGen}()}[R(\mathsf{SDec}(c), \mathsf{SDec}(c'')) = 1].$$

Observe that perfect non-malleability implies perfect security (but not vice versa).

The computational definition of non-malleability is more involved, using an indistinguishability experiment similar to that of the CCA-security definition. It formalizes the same idea, that an attacker must be unable (with more than negligible advantage) to modify ciphertexts such that the new decryption satisfies a known relation with the original decryption. The definition of non-malleability for (public-key) encryption schemes is based on the following experiment.

The NM indistinguishability experiment $\mathsf{PubK}^{\mathsf{NM}}_{\mathcal{A},\Pi}(k)$:

1. The challenger generates a key pair $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$ and sends $(1^k, pk)$ to $\mathcal{A}$.
2. $\mathcal{A}$ has oracle access to PDec_{sk}, and outputs (a description of) an efficiently samplable distribution M on the message space $\mathcal{M}$ (which must give non-zero probability only to strings of a given length).

3. The challenger samples a message $m \leftarrow M$, and sends ciphertext $c = \mathsf{PEnc}_{pk}(m)$ to $\mathcal{A}$.
4. $\mathcal{A}$ still has oracle access to PDec_{sk}, but cannot query $\mathsf{PDec}_{sk}(c)$. $\mathcal{A}$ outputs a ciphertext c' and (a description of) an efficiently computable relation $R : \mathcal{M} \times \mathcal{M} \to \{0, 1\}$.
5. The output of the experiment is 1 if $c' \neq c$ and $R(m, \mathsf{PDec}_{sk}(c'))$ is true, and 0 otherwise.

Define $\mathsf{PubK}^{\mathsf{NM},\$}_{\mathcal{A},\Pi}(k)$ to be identical to $\mathsf{PubK}^{\mathsf{NM}}_{\mathcal{A},\Pi}(k)$, except that item 3 is replaced by:

3′. The challenger samples independent messages $m, \tilde{m} \leftarrow M$, and sends $c = \mathsf{PEnc}_{pk}(\tilde{m})$ to $\mathcal{A}$.

Definition 3.6 (Computationally non-malleable encryption)**.** *A public-key encryption scheme* $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ *is* NM-CCA-secure *(that is, non-malleable against chosen ciphertext attacks), if for all* PPT *adversaries* $\mathcal{A}$ *there exists a negligible function* negl *such that*

$$\left|\Pr[\mathsf{PubK}^{\mathsf{NM}}_{\mathcal{A},\Pi}(k) = 1] - \Pr[\mathsf{PubK}^{\mathsf{NM},\$}_{\mathcal{A},\Pi}(k) = 1]\right| \leq \mathsf{negl}(k).$$

In our setting[2], CCA-security is equivalent to computational non-malleability, as stated in Claim 3.7. For the proof, we refer the reader to [BDPR98].

Claim 3.7. *An encryption scheme is* CCA*-secure (Definition 3.4) if and only if it satisfies computational non-malleability (Definition 3.6).*

3.4 Secure multi-party computation

Consider N players, each with an input value x_i for $i \in N$, who wish to jointly compute a function f on their inputs: $f(x_1, \cdots, x_N) = (y_1, \ldots, y_N)$. These players do not trust each other: they want each player i to receive his output value y_i at the end of the computation, but they also want a guarantee that no player i can learn any information beyond his designated output y_i (even if he "cheats"). Multi-party computation gives interactive N-party protocols to solve this problem, with security and correctness guarantees even when some players may maliciously deviate from the protocol.

Definition 3.8 (Secure multi-party computation)**.** *An N-party computation protocol is said to be* perfectly secure *(for up to $t < N$ corruptions) if it satisfies the following properties, against any adversary who corrupts up to t players[3]:*

- Correctness*: The output of the computation is equal to $f(x_1, \cdots, x_N)$.*
- Privacy*: No adversary can obtain any information about the honest parties' inputs, other than what can be deduced from the corrupted players' input and output values $\{x_i, y_i\}_{i \in S}$ (where S denotes the set of corrupt players).*

The protocol is said to be computationally secure *if it satisfies the above properties with all but negligible probability (in a security parameter k) against* PPT *adversaries.*

[2]When considering security notions other than CCA, standard indistinguishability-based security does not imply non-malleability. In this work we only use CCA-secure schemes.

[3]The corrupted players may be thought of as "dishonest" players trying to sabotage the protocol.

The following are general possibility results for multi-party computation that are relevant to this work. For proofs, we refer the reader to the original papers.[4]

Theorem 3.9 ([BGW88; CCD88])**.** *Any circuit can be evaluated by an N-party protocol with perfect security against $t < N/3$ corruptions. Moreover, the bound of $t < N/3$ is tight.*

Theorem 3.10 ([GMW87; AL11])**.** *Any circuit can be evaluated by an N-party protocol with computational security against up to $t = N - 1$ corruptions.*

An additional desirable property of multi-party computation protocols, other than correctness and privacy, is *guaranteed output delivery*: the property that every honest (non-corrupt) player is guaranteed to receive her correct output, even in the presence of an adversary. This property is known to be achievable if and only if $t < N/2$ (that is, a majority of the players are honest) [GMW87; Cle86].

3.5 Secret sharing

A secret sharing scheme specifies a method for a special party (the "dealer") to *share* a secret s among N players so that only large enough subsets of players can reconstruct the secret value s. The dealer gives privately a share s_i to each player i, so that any set of up to $t-1$ shares contains no information about s; however, it can efficiently be reconstructed given any t or more shares. The formal definition is given below.

Definition 3.11 (Secret sharing scheme [Sha79])**.** *A* t-out-of-N secret sharing scheme *is a pair of algorithms* $(\mathsf{Share}, \mathsf{Reconstruct})$ *as follows.* Share *takes as input a secret value s and outputs a set of* shares $S = \{s_1, \ldots, s_N\}$ *such that the following two properties hold.*

- Correctness*: For any subset $S' \subseteq S$ of size $|S'| \geq t$, it holds that $\mathsf{Reconstruct}(S') = s$, and*
- Privacy*: For any subset $S' \subseteq S$ of size $|S'| < t$, it holds that $H(s) = H(s|S')$, where H denotes the binary entropy function.*

$\mathsf{Reconstruct}$ *takes as input a (sub)set S' of shares and outputs:*

$$\mathsf{Reconstruct}(S') = \begin{cases} \perp & \text{if} \quad |S'| < t \\ s & \text{if} \quad \exists S \text{ s.t. } S' \subseteq S \text{ and } \mathsf{Share}(s) = S \text{ and } |S'| \geq t \end{cases}.$$

4 Cryptographically blinded games

Now we define "cryptographically blinded" games Γ' whose actions are encryptions of the actions of an underlying game Γ. Payoffs from corresponding action profiles of Γ and Γ' are the same. These blinded games will be an essential tool for our pre-play protocols, which will be detailed in Section 5.

The following supporting definition formalizes the intuitive notion that two strategic games are equivalent up to renaming of actions or deletion of redundant actions.

Definition 4.1. *For any strategic game $\Gamma = \langle N, (A_i), (u_i) \rangle$, a strategic game $\Gamma' = \langle N, (A'_i), (u'_i) \rangle$ is said to be* super-equivalent *to Γ if there exist surjective* renaming functions $\rho_i : A'_i \to A_i$ *such that for all $i \in N$, for all $a'_1 \in A'_1, \ldots, a'_N \in A'_N$, it holds that $u'_i(a'_1, \ldots, a'_N) = u_i(\rho_1(a'_1), \ldots, \rho_N(a'_N))$. In this case, we write $\Gamma' \geq_\rho \Gamma$.*

[4] Dodis and Rabin [DR07] provide and extended summary of the multi-party computation results with emphasis on the use in the game theoretical context.

Notation. For a renaming function ρ, let $\rho_i^{-1} : A_i \to \mathcal{P}(A_i')$ be defined by $\rho_i^{-1}(a_i) = \{a_i' | \rho(a_i') = a_i\}$. To simplify notation, we define $\rho : A_1 \times \cdots \times A_N \to A_1' \times \cdots \times A_N'$ to be $\rho(a_1, \ldots, a_N) = (\rho_1(a_1), \ldots, \rho_N(a_N))$, and let ρ^{-1} be defined similarly. For a distribution γ' on action profiles of Γ', $\rho(\gamma')$ denotes the distribution on action profiles of Γ that corresponds to sampling $a' \in A'$ according to γ' and outputting $\rho(a')$.

Lemma 4.2. *Let Γ be a strategic game. Then for any Γ' with $\Gamma' \geq_\rho \Gamma$ it holds that: (1) for any coarse correlated equilibrium α of Γ, there exists a coarse correlated equilibrium α' of Γ' such that $\rho(\alpha') = \alpha$; and (2) for any coarse correlated equilibrium α' of Γ', $\rho(\alpha')$ is a coarse correlated equilibrium of Γ.*

Proof. To show item (1), consider the distribution α' on action profiles of Γ' obtained by sampling an action profile a from α and outputting a random $a' \in \rho^{-1}(a)$. Note that $\rho(\alpha') = \alpha$ by construction. We need to show that for all $i \in N$ and all $a_i^* \in A_i'$,

$$\mathop{\mathbf{E}}_{a' \leftarrow \alpha'}[u_i'(a')] \geq \mathop{\mathbf{E}}_{a' \leftarrow \alpha'}[u_i'(a_i^*, a_{-i}')].$$

The above can be rewritten, due to the construction of α' and definition of Γ', as $\mathbf{E}_{a \leftarrow \rho(\alpha')}[u_i(a)] \geq \mathbf{E}_{a \leftarrow \rho(\alpha')}[u_i(\rho_i(a_i^*), a_{-i})]$. This holds for every $\rho_i(a_i^*) \in A_i$ since $\alpha = \rho(\alpha')$ is a coarse correlated equilibrium of Γ. Item (2) follows similarly, since $\rho(\alpha')$ is a distribution on action profiles of Γ and α' is a coarse correlated equilibrium. □

The interesting case of the seemingly straightforward definition of super-equivalence arises when the renaming function ρ is not invertible by the players.

We now define cryptographically blinded games. Let $\Gamma = \langle N, (A_i), (u_i) \rangle$ be a strategic game, where players have oracle access to the utility functions u_i.

Definition 4.3 (Secret-key blinded game)**.** *Let $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ be a secret-key encryption scheme, and let $\Gamma = \langle N, (A_i), (u_i) \rangle$ be a strategic game. Define the* blinded game $\Gamma^\Sigma = \langle N, (A_i'), (u_i)' \rangle$ *of Γ to be the game such that $sk \leftarrow \mathsf{SGen}()$ is generated and*

- *for each player $i \in N$ the action space is $A_i' = A_i \sqcup \{\mathsf{SEnc}_{sk}(a_i) | a_i \in A_i\}$*
- *for each player $i \in N$ the utility for all $a' \in \times_{j \in N} A_j'$ is $u_i'(a') = u_i(a)$, where for all $j \in N$*

$$a_j = \begin{cases} a_j' & \text{if } a_j' \in A_j, \\ \mathsf{SDec}_{sk}(a_j') & \text{otherwise.} \end{cases}$$

Definition 4.4 (Public-key blinded game)**.** *Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a public-key encryption scheme, and let $\Gamma = \langle N, (A_i), (u_i) \rangle$ be a strategic game. Define the* computational blinded game $\Gamma^\Pi = \{\langle N, (A_i'^{(k)}), (u_i'^{(k)}) \rangle\}_{k \in \mathbb{N}}$ *of Γ to be the computational game such that for every security parameter $k \in \mathbb{N}$ a corresponding key pair $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$ is generated and*

- *for each player $i \in N$ the action space is $A_i'^{(k)} = \{\mathsf{PEnc}_{pk}(a_i) | a_i \in A_i\}$*
- *for each player $i \in N$ the utility for all $a' \in \times_{i \in N} A_i'^{(k)}$ is $u_i'^{(k)}(a') = u_i(\mathsf{PDec}_{sk}(a'))$.*

If Γ^Π and Γ^Σ are blinded games of the game Γ, then we say that Γ is the *underlying game* of Γ^Π and Γ^Σ.

Observe that the blinded games Γ^Π and Γ^Σ are super-equivalent to the underlying game Γ, with respect to renaming functions $\rho = \mathsf{PDec}_{sk}$ or $\rho = \mathsf{SDec}_{sk}$ (respectively).

Remark. In these contexts, players do not have knowledge of the secret key sk, as is standard and necessary when employing encryption schemes. Therefore, expectations "from the point of view of the player" are taken over a secret key $sk \leftarrow \mathsf{SGen}()$ or $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$, where secret- or public-key encryption schemes are used, respectively.

It is assumed to be infeasible for players of a game Γ to efficiently compute the utility functions u_i' on arbitrary action profiles in Γ^Σ or Γ^Π, since they cannot (efficiently) decrypt ciphertexts in the corresponding encryption schemes. However, our applications require players to be able to pick actions in A_i' for which they know the corresponding expected utility. In fact, if the players cannot do this, then the games become meaningless in that *any* distribution on A is an equilibrium. In the public-key case, this property is achieved as players can simply compute the encryption of some $a_i \in A_i$ for which the utility is known. In the secret-key case, A_i is contained in A_i' for exactly this purpose.

Security parameter for public-key games. Public-key blinded games have an implicit security parameter k due to the underlying encryption scheme. When applying computational equilibrium concepts (which have a security parameter k' of their own) to such games, there must be a fixed relation between k and k' in order to have a meaningful definition of security for a computational equilibrium of a blinded game. In our setting, both parameters represent the same quantity: the computational boundedness of the players of a game. Therefore, we let $k = k'$ and refer to a single security parameter k.

4.1 Correspondence of equilibria in blinded games

Lemma 4.5. *Let $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ be a perfectly non-malleable and verifiably decryptable secret-key encryption scheme. Then for any strategic game Γ, it holds that for any coarse correlated equilibrium α of Γ there exists a correlated equilibrium α' of Γ^Σ that achieves the same utility profile as α.*

Proof. Let α' be the probability distribution on $\times_{i \in N} A_i'$ that corresponds to sampling an action profile $a = (a_1, \ldots, a_N) \in \times_{i \in N} A_i$ according to α and outputting an action profile $a' = (\mathsf{SEnc}_{sk}(a_1), \ldots, \mathsf{SEnc}_{sk}(a_N))$, where sk is the secret key generated by SGen. Note that α' achieves the same utility profile as α by construction.

To show that such α' constitutes a correlated equilibrium of Γ^Σ, we need to verify that the conditions from Definition 2.3 are satisfied, i.e. for every player i and for all $b_i', \hat{a}_i' \in A_i'$ it must hold that

$$\mathop{\mathbf{E}}_{sk \leftarrow \mathsf{SGen}(), a' \leftarrow \alpha'} [u_i'(a') | a_i' = b_i'] \geq \mathop{\mathbf{E}}_{sk \leftarrow \mathsf{SGen}(), a' \leftarrow \alpha'} [u_i'(\hat{a}_i', a'_{-i}) | a_i' = b_i']. \tag{1}$$

Since Σ is perfectly secure, it follows from Definition 3.3 that for any $a_0', a_1' \in A_i'$,

$$\mathop{\mathbf{E}}_{sk \leftarrow \mathsf{SGen}(), a' \leftarrow \alpha'} [u_i'(a') | a_i' = a_0'] = \mathop{\mathbf{E}}_{sk \leftarrow \mathsf{SGen}(), a' \leftarrow \alpha'} [u_i'(a) | a_i' = a_1'].$$

Thus, for any player i, the expected utility from the distribution α' is independent of the advice a_i'. Moreover, since the underlying encryption scheme is perfectly non-malleable (Definition 3.5), no player i can generate (with any advantage[5]) a deviation a_i^* satisfying $R(a_i^*, a_i)$ for any known relation R. It follows that we need only to consider deviations a_i^* that are independent of the received advice a_i. Therefore, equation 1 can be rewritten as the following: for every player i and for all $\hat{a}_i' \in A_i'$ independent of a_i',

$$\mathop{\mathbf{E}}_{sk \leftarrow \mathsf{SGen}(), a' \leftarrow \alpha'} [u_i'(a')] \geq \mathop{\mathbf{E}}_{sk \leftarrow \mathsf{SGen}(), a' \leftarrow \alpha'} [u_i'(\hat{a}_i', a'_{-i})],$$

[5]More precisely, no player can generate such a deviation a_i^* with more success than by random guessing.

which holds because α' is by Lemma 4.2 a coarse correlated equilibrium of Γ^{Σ}. $\square$

Lemma 4.6. *Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a CCA-secure public-key encryption scheme. Then for any strategic game Γ, it holds that for any computational coarse correlated equilibrium α of Γ there exists a computational correlated equilibrium α' of Γ^{Π} that achieves the same utility profile as α.*

Proof. For each security parameter $k \in \mathbb{N}$, let (pk, sk) be the corresponding key pair generated by $\mathsf{PGen}(1^k)$. Consider the following probability ensemble $\alpha' = \{\alpha'^{(k)}\}_{k\in\mathbb{N}}$ on $\{\times_{j\in N} A_j'^{(k)}\}_{k\in\mathbb{N}}$ that corresponds for each $k \in \mathbb{N}$ to sampling an action profile $a = (a_1, \ldots, a_N) \in \times_{i\in N} A_i$ according to $\alpha^{(k)}$ and outputting an action profile $a' = (\mathsf{PEnc}_{pk}(a_1), \ldots, \mathsf{PEnc}_{pk}(a_N))$. Note that α' achieves the same utility profile as α by construction.

Assume that α' is not a computational correlated equilibrium of Γ^{Π} (Definition 2.6), i.e. there exist a player $i \in N$, a PPT-samplable ensemble $\hat{\alpha}_i' = \{\hat{\alpha}_i'^{(k)}\}_{k\in\mathbb{N}}$ on $\{A_i'^{(k)}\}_{k\in\mathbb{N}}$, and a non-negligible function $\delta(\cdot)$ such that for every $k \in \mathbb{N}$

$$\mathop{\mathbf{E}}_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k),\\ a'\leftarrow\alpha'^{(k)}}}[u_i'^{(k)}(a')] \leq \mathop{\mathbf{E}}_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k),\\ a'\leftarrow\alpha'^{(k)},\hat{a}_i'\leftarrow\hat{\alpha}_i'^{(k)}(a_i')}}[u_i'^{(k)}(\hat{a}_i', a'_{-i})] - \delta(k). \tag{2}$$

We show that one can use such a deviation $\hat{\alpha}_i'$ to construct a PPT adversary that contradicts the computational non-malleability of the encryption scheme Π (Definition 3.6).

Let $\mathcal{A}$ be the adversary that for each security parameter $k \in \mathbb{N}$ behaves as follows. $\mathcal{A}$ receives a public key pk from the challenger and sends back $M = \alpha^{(k)}$ as the message distribution. Upon receiving the challenge ciphertext c the adversary $\mathcal{A}$ samples $c' \leftarrow \hat{\alpha}_i'^{(k)}(c)$ and sends c' to the challenger together with the relation

$$R(b, \hat{b}) = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \cdot (\mathbf{E}_{a\leftarrow\alpha^{(k)}}[u_i(\hat{b}, a_{-i})|a_i = b] - \mathbf{E}_{a\leftarrow\alpha^{(k)}}[u_i(a_i, a_{-i})|a_i = b] + 1), \\ 0 & \text{otherwise.} \end{cases}$$

We can assume without loss of generality that all the utilities of all the players in Γ are between 0 and 1 (the corresponding linear transformation of the game matrix does not change the strategic properties of the game), hence the above expression defining the probability that $R(b, \hat{b})$ holds is between 0 and 1. Note that M is efficiently samplable and that the relation R is efficiently computable.

Consider the success probability of $\mathcal{A}$ in the experiment $\mathsf{PubK}^{\mathsf{NM}}_{\mathcal{A},\Pi}(k)$, i.e.

$$\begin{aligned} \Pr[\mathsf{PubK}^{\mathsf{NM}}_{\mathcal{A},\Pi}(k) = 1] &= \Pr_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k)\\ a\leftarrow\alpha^{(k)},\hat{a}_i'\leftarrow\hat{\alpha}_i'^{(k)}(\mathsf{PEnc}_{sk}(a_i))}}[\hat{a}_i' \neq \mathsf{PEnc}_{sk}(a_i) \wedge R(a_i, \mathsf{PDec}_{sk}(\hat{a}_i'))] \\ &= \frac{1}{2}\left(\mathop{\mathbf{E}}_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k),\\ a'\leftarrow\alpha'^{(k)},\hat{a}_i'\leftarrow\hat{\alpha}_i'^{(k)}(a_i')}}[u_i'^{(k)}(\hat{a}_i', a'_{-i})] - \mathop{\mathbf{E}}_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k),\\ a'\leftarrow\alpha'^{(k)}}}[u_i'^{(k)}(a')] + 1 \right). \end{aligned}$$

Note that the scaling needed for relation R is done by some finite factor, since the game matrix of Γ does not depend on the security parameter k. Therefore, it follows from equation 2 that this probability is larger than $\delta'(k)$ for some non-negligible function $\delta'(\cdot)$.

On the other hand, the success probability of $\mathcal{A}$ in the experiment $\mathsf{PubK}^{\mathsf{NM},\$}_{\mathcal{A},\Pi}(k)$, i.e.

$$\begin{aligned}\Pr[\mathsf{PubK}^{\mathsf{NM},\$}_{\mathcal{A},\Pi}(k)=1] &= \Pr_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k)\\ a,\tilde{a}\leftarrow\alpha^{(k)},\hat{a}'_i\leftarrow\hat{\alpha}_i'^{(k)}(\mathsf{PEnc}_{sk}(a_i))}}[\hat{a}'_i\neq\mathsf{PEnc}_{sk}(a_i)\wedge R(\tilde{a}_i,\mathsf{PDec}_{sk}(\hat{a}'_i))]\\ &= \frac{1}{2}\Bigg(\mathop{\mathbf{E}}_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k),\\ a'\leftarrow\alpha'^{(k)},\hat{a}'_i\leftarrow\hat{\alpha}_i'^{(k)}}}[u_i'^{(k)}(\hat{a}'_i,a'_{-i})]-\mathop{\mathbf{E}}_{\substack{(pk,sk)\leftarrow\mathsf{PGen}(1^k),\\ a'\leftarrow\alpha'^{(k)}}}[u_i'^{(k)}(a')]+1\Bigg),\end{aligned}$$

can be at most $\epsilon(k)$ for some negligible function ϵ. This follows from the fact that α is a computational coarse correlated equilibrium, and no independent deviation can yield a non-negligible improvement in expectation on the utility of any player i.

Putting the above two observations together we conclude that for some non-negligible $\delta^*(\cdot)$

$$\left|\Pr[\mathsf{PubK}^{\mathsf{NM}}_{\mathcal{A},\Pi}(k)=1]-\Pr[\mathsf{PubK}^{\mathsf{NM},\$}_{\mathcal{A},\Pi}(k)=1]\right|\geq\delta^*(k),$$

a contradiction to computational non-malleability of Π. □

4.2 What can I do with an encrypted action?

We employ blinded games as a tool to achieve equilibria in the underlying game. The pre-play protocols in the next section will issue "advice" to the players as *encrypted actions*, that is, actions in the blinded game. In this section we address how an action of the blinded game can be "used" to take a corresponding action in the underlying game.

We return to the concept of verifiability of mediation, introduced in Section 1. Since the players do not know the secret key associated with a blinded game, they cannot decrypt an encrypted action (and indeed, this is an essential property upon which the pre-play protocols will depend). The players therefore invoke a third party who plays the underlying game *on their behalf*. The third party will act in a way which can be publicly verified, so no trust need be placed in him to perform actions correctly: if he misbehaves, then the misconduct will be detected and he can be held accountable. This is in contrast to the usual idea of trusted mediation for implementation of equilibria.

The importance of reducing the trust placed in mediators has long been recognized in the literature, and the first formal definition of a verifiable but not trusted form of mediation was given in [ILM11], which introduced the concept of *verifiable mediator*.

Definition 4.7 (Verifiable mediator [ILM11])**.** *A* verifiable mediator *is a mediator which performs all actions in a publicly verifiable way, and does not use any information that must be kept secret.*

We introduce the new concept of a *verifiable proxy*, which is used in our construction. Note that the new concept is incomparable to the verifiable mediator of [ILM11].

Definition 4.8 (Verifiable proxy)**.** *A* verifiable proxy *is a mediator which performs all actions in a publicly verifiable way, and does not give the players any information that affects their strategic choices in the underlying strategic game.*

In our setting, the (only) action that the verifiable proxy performs for the players is to *translate* the action from an encrypted form to the original form. It is well known that decryption can be done verifiably (see Appendix B for details). Importantly, the verifiable proxy acts independently for each player: the correlation between players' strategies is achieved by the

players themselves with no external help, and the verifiable proxy acts simply as a *proxy* or interface so that the players may use encrypted actions to play in the underlying game.

We believe that (in contrast to general trusted mediators), verifiable proxies are a very realistic and mild requirement in many scenarios, since many games are already "set up" by some entity (e.g. the stock exchange or an online games company), which could easily set up instead a version of the game incorporating encrypted actions. Moreover, the impossibility result of [HNR13] shows that without any mediation, even correlated equilibria cannot in general be achieved by cheap talk: so some weak notion of mediation is necessary in order to bypass this result and give useful correlated equilibrium implementations.

Example 4.9. More concretely, we provide a toy example involving the well-known "Battle of the Sexes" game (Figure 1), where two friends are deciding on a joint activity, and they have opposing preferences but would rather be together than apart:

	Bach (B)	Stravinsky (S)
Bach (B)	$\mathbf{2}, \mathbf{5}$	$0, 0$
Stravinsky (S)	$0, 0$	$\mathbf{5}, \mathbf{2}$

Figure 1: "Battle of the Sexes" game

It is a correlated equilibrium to randomize over (B, B) and (S, S). In this scenario, the "encrypted advice" could be an order to an online ticket vendor for either a Bach or Stravinsky concert, encrypted under the public key of the vendor. The set-up assumption here would be that the online vendor has published a public key and accepts encrypted orders. Since accepting orders in a variety of formats desirable to customers is in the vendor's interest, we consider this to be a very feasible scenario.

Note that as this particular example is a correlated equilibrium, it is unnecessary to encrypt advice (e.g. since the protocol of [DHR00] applies). However, the example serves to illustrate that verifiable translation can be a highly realistic and mild assumption.

5 Our Protocols

In this section we give cryptographic protocols (in the computational and information-theoretic settings) that achieve the utility profile of any coarse correlated equilibrium.

5.1 Cryptographic cheap talk

Definition 5.1 (Cheap talk extension)**.** *For a strategic game* Γ*, the* cheap talk extension $\widetilde{\Gamma}$ *is defined as an extensive game consisting of a pre-play phase in which the players exchange messages, followed by the play in the original strategic game. The communication is non-binding (unlike in signaling games) in that it does not directly affect players' utilities in the underlying game, that is, players' utilities in the cheap talk extension depend only on actions taken in the strategic game. The* cryptographic cheap talk extension *is defined exactly like the cheap talk extension, except that the players exchange messages during a polynomially bounded number of rounds prior to the play in the original game* Γ*.*

We follow the pre-play paradigm of [Bár92], where the mediator is replaced by "cheap talk" communication prior to game play. We construct protocols to be run during pre-play, which implement any (computational) coarse correlated equilibrium of blinded games as a (computational) Nash equilibrium of the (computational) cheap talk extension.

5.2 Protocol for computationally bounded players

In this protocol, the players run a computationally secure multi-party computation to sample an action profile from any computational correlated equilibrium of the blinded game.

Protocol 1. Implementing any computational correlated equilibrium α' of Γ^{Π}:

Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a CCA-secure public-key encryption scheme and let $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$ with pk known to all players. Communication is via broadcast.

1. The players run a computationally secure multi-party computation protocol (secure against $t \leq N-1$ corruptions) to implement the function that samples an action profile $a' \leftarrow \alpha'$, and outputs to each player i his action a'_i.
2. Every player takes a'_i as its action in Γ^{Π}.

We show that rational computationally bounded players will follow the above protocol, so they can use it to implement any computational correlated equilibrium. Then, by combining the above with our results from Section 4 about correspondence of coarse correlated equilibria in the underlying game and correlated equilibria in its blinded version, we obtain that the protocol can moreover be used to implement any computational *coarse* correlated equilibrium.

Note that it is necessary to treat the two-player case somewhat differently from the case with three or more players, because of the problem of guaranteed output delivery in the two-player case (which was described in Section 3.4). We begin by presenting the simpler Theorem 5.2, which states that Protocol 1 can be *directly* used by three or more players to implement any computational coarse correlated equilibrium. Then, we give Theorems 5.3 and 5.4 which show that by running a *slightly modified* version of Protocol 1, it is possible for *any* number of players to implement any computational coarse correlated equilibrium.

Theorem 5.2. *Let* $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ *be a* CCA*-secure public-key encryption scheme, and let* Γ *be any finite strategic game with three or more players. For any computational coarse correlated equilibrium* α *of* Γ*, there exists a computational Nash equilibrium* $\widetilde{\alpha}$ *of the computational cheap talk extension* $\widetilde{\Gamma^{\Pi}}$ *that achieves the same utility profile as* α*.*

Proof. Let α' be the computational correlated equilibrium of Γ^{Π} from Lemma 4.6 that achieves the same utility profile as α. We show that using Protocol 1 in order to implement α' constitutes a computational Nash equilibrium in the cryptographic cheap talk extension $\widetilde{\Gamma^{\Pi}}$. Note that it is payoff-equivalent to α by construction.

By the *privacy* guarantee of the secure multi-party computation protocol, we have that no player can learn any (non-negligible amount of) information that cannot be deduced from his intended output in the first place, even if he deviates from the protocol arbitrarily. Moreover, since there are three or more players and we consider only unilateral[6] deviations (as implied by the definition of Nash equilibrium), the protocol has the property of *guaranteed output delivery*[7]:

[6] That is, we only consider deviations from the protocol by a single (malicious) player, rather than by coalitions of multiple colluding players.

[7] We remark that in fact, the slightly weaker property of *fairness* is sufficient: that is, the property that if any player receives his output in the protocol, then every honest player will receive her correct output too. However, in the settings we consider, the stronger property of *guaranteed output delivery* is known to hold, hence we refer to the latter property in order to slightly simplify the proof.

therefore, the deviation of any player i cannot prevent any other player j from receiving her correct output a'_j.

We have shown that for any player, there is no deviation during the protocol phase that is profitable by more than negligible amount. Hence, we consider only the case where each player i receives his correct output a'_i. Since α' is, by Lemma 4.6, a computational correlated equilibrium of Γ^{Π}, no player has an incentive to deviate from the prescribed advice, and thus the players will play according to the sampled action profile a'. Therefore, to follow Protocol 1 is the computational Nash equilibrium $\widetilde{\alpha}$ of $\widetilde{\Gamma^{\Pi}}$ payoff-equivalent to α. □

5.2.1 Dealing with the two-player case

In the two-player case, the additional complication stems from the fact that in this setting we do not have guaranteed output delivery: hence, it is necessary to consider that a player may be incentivized to cause a protocol execution to terminate prematurely. In order to disincentivize such behavior, we introduce an additional "punishment" condition to the protocol, as follows.

Protocol 2. Implementing any computational correlated equilibrium α' of Γ^{Π}:

Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a CCA-secure public-key encryption scheme and let $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$ with pk known to all players. Communication is via broadcast.

- The players run Protocol 1 as long as no player is detected to deviate from the protocol.
- If any player i is detected to deviate from the protocol, then all (other) players adopt the strategies (in Γ^{Π}) corresponding to the worst Nash equilibrium σ^i for player i.

Using Protocol 2, we obtain the following theorem that applies for *any* number of players.

Theorem 5.3. *Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a CCA-secure public-key encryption scheme, and let Γ be any finite strategic game. For any computational coarse correlated equilibrium α of Γ that for each player achieves at least as high utility as the worst Nash equilibrium, there exists a computational Nash equilibrium $\widetilde{\alpha}$ of the computational cheap talk extension $\widetilde{\Gamma^{\Pi}}$ that achieves the same utility profile as α.*

Proof. Let α' be the computational correlated equilibrium of Γ^{Π} from Lemma 4.6 that achieves the same utility profile as α. We show that using Protocol 2 in order to implement α' constitutes a computational Nash equilibrium in the cryptographic cheap talk extension $\widetilde{\Gamma^{\Pi}}$. For any security parameter k, the following events may occur during the run of the protocol:

1. a player learns its advice before the other players;
2. a player deviates from the protocol and the deviation is detected by the other players; or
3. a player deviates from the protocol and it is unnoticed.

Addressing (1): it follows from CCA-security of the public-key encryption scheme Π (Definition 3.4) that each player is indifferent (up to a negligible improvement in utility) between any advice he may receive, and thus gains no advantage from learning his advice first. In particular, he has no incentive to abort the protocol and prevent others from learning their advice. Addressing (2): the expectation of any player i in the default Nash equilibrium σ^i is at most the

expectation of player i in α. Addressing (3): the security of the multi-party computation protocol ensures that players can cheat without being caught with at most negligible probability. Thus, the increase in utility from any cheating strategy is at most negligible.

There is no deviation during the protocol phase profitable by more than negligible amount. Consider the case that every player i received his advice a'_i. Since α' is, by Lemma 4.6, a computational correlated equilibrium of Γ^{Π}, no player has an incentive to deviate from the prescribed advice, and the players will play according to the sampled action profile a'. Therefore, to follow Protocol 2 is the computational Nash equilibrium $\widetilde{\alpha}$ of $\widetilde{\Gamma^{\Pi}}$ payoff-equivalent to α. □

It is possible to eliminate the condition (from Theorem 5.3) that the implemented coarse correlated equilibrium does at least as well as the respective worst Nash equilibrium for each player, thereby obtaining a yet more general theorem as follows.

Theorem 5.4. *Let* $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ *be a* CCA*-secure public-key encryption scheme, and let* Γ *be any finite strategic game. For any coarse correlated equilibrium* α *of* Γ*, there exists a computational Nash equilibrium* $\widetilde{\alpha}$ *of the computational cheap talk extension* $\widetilde{\Gamma^{\Pi}}$ *that achieves the same utility profile as* α*.*

The proof of Theorem 5.4 makes use of another variant of Protocol 1. The details of this variant protocol (Protocol 4) are given in Appendix C along with the proof of the theorem.

We remark that Protocol 2 has certain more desirable properties than Protocol 4: in particular, Protocol 2 is *free of empty threats*, which ensures that Nash equilibria in the protocol are stable even when players may change strategy *adaptively* during protocol execution (a formal definition of empty threats may be found in Appendix D). Ultimately, notwithstanding the restriction on the class of achieved coarse correlated equilibria, we consider Theorem 5.3 to be the much stronger result compared to Theorem 5.4, for the following reasons:

- all coarse correlated equilibria that players might rationally wish to implement by cheap talk do dominate all Nash equilibria (otherwise, they could achieve a better payoff from a Nash equilibrium without the hassle of a pre-play protocol); and
- unlike Protocol 4, Protocol 2 is free of empty threats; and
- the expected payoff even when the protocol is aborted and the default strategy invoked is higher in Protocol 2 than in Protocol 4.

Strategic equivalence. Lemma 5.5, below, proves the strategic equivalence of the cryptographic cheap talk extension $\widetilde{\Gamma^{\Pi}}$ to the underlying game Γ.

Lemma 5.5. *Let* $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ *be a* CCA*-secure public-key encryption scheme, and let* Γ *be any finite strategic game. For any computational Nash equilibrium* $\widetilde{\alpha}$ *of the cryptographic cheap talk extension* $\widetilde{\Gamma^{\Pi}}$*, there exists a computational coarse correlated equilibrium* α *of* Γ *that achieves the same utility profile as* $\widetilde{\alpha}$*.*

Proof. We show that the probability ensemble α induced by $\widetilde{\alpha}$ on action profiles of Γ is a computational coarse correlated equilibrium of Γ.

Assume that α is not a computational coarse correlated equilibrium, i.e. there exists a player i that has a PPT-samplable unilateral deviation to α that improves his expectation for every $k \in \mathbb{N}$ by $\delta(k)$ for some non-negligible $\delta(\cdot)$. However, such deviation can be used by player i also against $\widetilde{\alpha}$ to gain a non-negligible improvement in his expectation in $\widetilde{\Gamma^{\Pi}}$, a contradiction to the fact that $\widetilde{\alpha}$ is a computational Nash equilibrium of $\widetilde{\Gamma^{\Pi}}$. □

Corollary 5.6. *For any finite strategic game Γ, the cryptographic cheap talk extension $\widetilde{\Gamma^{\Pi}}$ is strategically equivalent to Γ, that is, for every Nash equilibrium $\widetilde{\alpha}$ of $\widetilde{\Gamma^{\Sigma}}$, there exists a coarse correlated equilibrium of Γ that achieves the same utility profile as $\widetilde{\alpha}$, and vice versa.*

Proof. Follows immediately from Lemma 5.5 and Theorem 5.3 (or Theorem 5.2 for the case of three or more players). □

5.3 Protocol for computationally unbounded players

An alternative protocol using secret-key encryption implements all coarse correlated equilibria – not just computational ones – for all strategic games with four or more players. As discussed in Section 1, the condition of four or more players is unavoidable. In this (more traditional) setting, the players are computationally unbounded.

Protocol 3. Implementing any correlated equilibrium α' of Γ^{Σ}:

Let $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ be a perfectly non-malleable and verifiably decryptable secret-key encryption scheme and let $sk \leftarrow \mathsf{SGen}$. Let each player i possess a distinct share sk_i of an $(N-1)$-out-of-N secret-sharing $\{sk_1, \ldots, sk_N\}$ of sk. Communication is via pairwise channels.

1. The players run a perfectly secure multi-party computation to implement the function that samples a profile $a' \leftarrow \alpha'$, and outputs to each i his action a'_i.
2. Every player takes a'_i as its action in Γ^{Σ}.

Theorem 5.7. *Let $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ be a perfectly non-malleable and verifiably decryptable secret-key encryption scheme, and let Γ be any finite strategic game with four or more players. For any coarse correlated equilibrium α of Γ there exists a Nash equilibrium $\widetilde{\alpha}$ of the cheap talk extension $\widetilde{\Gamma^{\Sigma}}$ that achieves the same utility profile as α.*

Proof. Let α' be the correlated equilibrium of Γ^{Σ} from Lemma 4.5 that achieves the same utility profile as α. We show that to follow Protocol 3 in order to implement α' constitutes the Nash equilibrium $\widetilde{\alpha}$ in the cryptographic cheap talk extension $\widetilde{\Gamma^{\Sigma}}$ that achieves the same utility profile as α.

First note that since the players are using a perfectly secure protocol with output guarantee (see Section 3.4) to implement sampling from α', no player can prevent the others from learning their advice by a unilateral deviation during the multi-party computation phase. Moreover, even if a single player i withholds its share sk_i the remaining players hold $N-1$ shares of the secret key sk that are sufficient to reconstruct the secret key and sample an action profile from α'. Hence, any unilateral deviation does not influence the distribution on actions taken by the other players. Assume that there exists a unilateral deviation for some player i in $\widetilde{\Gamma^{\Sigma}}$ that allows him to gain a higher utility than by playing according to $\widetilde{\alpha}$. This contradicts α' being a correlated equilibrium of Γ^{Σ}, since it could be used as a unilateral profitable deviation against α' in Γ^{Σ} as well. □

Strategic equivalence. Lemma 5.8, below, proves the strategic equivalence of the cheap talk extension $\widetilde{\Gamma^{\Sigma}}$ to the underlying game Γ.

Lemma 5.8. *Let $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ be a perfectly non-malleable and verifiably decryptable secret-key encryption scheme, and let Γ be any finite strategic game with four or more players. For any Nash equilibrium $\widetilde{\alpha}$ of the cheap talk extension $\widetilde{\Gamma^{\Sigma}}$, there exists a coarse correlated equilibrium α of Γ that achieves the same utility profile as $\widetilde{\alpha}$.*

Proof. We show that the distribution α induced by $\widetilde{\alpha}$ on action profiles of Γ is a coarse correlated equilibrium of Γ. Suppose α is not a coarse correlated equilibrium, i.e. there exists a player i that has a deviation to α which improves his expectation. However, such a deviation contradicts the fact that $\widetilde{\alpha}$ is a Nash equilibrium of $\widetilde{\Gamma^{\Sigma}}$, since it is also a profitable unilateral deviation against $\widetilde{\alpha}$ in $\widetilde{\Gamma^{\Sigma}}$. □

Corollary 5.9. *For any game Γ, it holds that the cheap talk extension $\widetilde{\Gamma^{\Sigma}}$ is strategically equivalent to Γ, that is, for every Nash equilibrium $\widetilde{\alpha}$ of $\widetilde{\Gamma^{\Sigma}}$, there exists a coarse correlated equilibrium of Γ that achieves the same utility profile as $\widetilde{\alpha}$, and vice versa.*

Proof. Follows immediately from Theorem 5.7 and Lemma 5.8. □

Sequential equilibrium. We also show that the equilibrium from Theorem 5.7 is a *sequential equilibrium* (relevant formal definitions are given in Appendix A): informally, we show that by following the prescribed strategy, the players are making optimal decisions at all points in the game tree. Our proof relies on perfect security for multi-party computation protocols in the presence of one actively corrupted and one passively corrupted party which can be achieved only for six or more players (as shown by Fitzi, Hirt and Maurer [FHM98], see Section 3.4). Hence, the statement of the following theorem is less general than the statement of Theorem 5.7.

Theorem 5.10. *Let $\Sigma = (\mathsf{SGen}, \mathsf{SEnc}, \mathsf{SDec})$ be a perfectly non-malleable and verifiably decryptable secret-key encryption scheme, and let Γ be any finite strategic game with six or more players. For any coarse correlated equilibrium α of Γ there exists a sequential equilibrium $(\widetilde{\alpha}, \mu)$ of the cheap talk extension $\widetilde{\Gamma^{\Sigma}}$ that achieves the same utility profile as α.*

Proof. We assume without loss of generality that the multi-party computation protocol has the canonical structure where at each round a single player receives a message from one of the other players (i.e. the information sets in the extensive game correspond to histories consistent with the received message). Since there is at least six players, we can assume that multi-party computation is secure in the presence of one static and one active corruption. Consider the behavioral strategy profile $\widetilde{\alpha}$ corresponding to following Protocol 3 at each history where a player receives a message from some other player (in particular this corresponds to ignoring all received messages after termination of the multi-party computation).

First, we specify the belief system μ of players at any information set. The beliefs at any information set on the equilibrium path are derived from the behavioral strategy $\widetilde{\alpha}$ by Bayes' rule, and for any information set I that lies off the equilibrium path (i.e. an information set corresponding to receiving a message out of the scope of the protocol), let $\mu(I)$ be the uniform distribution on all histories in I. To show that $(\widetilde{\alpha}, \mu)$ is a sequential equilibrium, we must show that $(\widetilde{\alpha}, \mu)$ is both sequentially rational and consistent.

Since $\widetilde{\alpha}$ is a Nash equilibrium (as shown in Theorem 5.7), the behavioral strategy to follow $\widetilde{\alpha}$ is optimal for any information set on the equilibrium path. Hence, to conclude that $(\widetilde{\alpha}, \mu)$ is sequentially rational, we just need to show that $\widetilde{\alpha}$ is also optimal off the equilibrium path, given the beliefs of μ. Let I be an information set of player i at some point off the equilibrium

path that corresponds to receiving a message from player j. Note that even if j sends to i its complete view of the protocol up to this point player i cannot use such information to produce a profitable deviation, since such deviation would imply an adversary corrupting actively player i and statically player j able to break the perfect security of the multi-party computation protocol. Now consider any history off the equilibrium path after the termination of the multi-party computation, and assume that player i receives the private advice of some other player. There cannot exist a profitable deviation of player i, since such a deviation would contradict security of the secret key encryption scheme.

To show that (β, μ) is consistent we use the "trembling-hand" approach. Consider the sequence of assessments $\{(\beta^{(n)}, \mu^{(n)})\}_{n=1}^{\infty}$ where each $\beta^{(n)}$ assigns non-zero probability $\epsilon^{(n)}$ to all actions that are taken with zero probability in β, such that $\epsilon^{(n)}$ goes to zero as $n \to \infty$, and the belief system $\mu^{(n)}$ is derived from $\beta^{(n)}$ using Bayes' rule. First note that the sequence $\{(\beta^{(n)}, \mu^{(n)})\}_{n=1}^{\infty}$ converges to (β, μ). The sequence of behavioral strategy profiles $\{\beta^{(n)}\}_{n=1}^{\infty}$ converges to β by construction. Since $\mu^{(n)}$ is derived from $\beta^{(n)}$ by the Bayes' rule, $\mu^{(n)}$ converges to μ for every information set on the equilibrium path. For every information set I off the equilibrium path, the distribution $\mu^{(n)}(I)$ is equal to $\mu(I)$. Finally, $\beta^{(n)}$ is completely mixed for all n, hence (β, μ) is consistent. □

5.4 Remarks on efficiency of multi-party computation

Computational setting. With recent advances in efficiency, computationally secure multi-party computation protocols are now being considered for practical use in various settings. Its first large-scale deployment was to compute market clearing prices for Danish sugar beet contracts in 2008 [Bog+09]. Subsequent advances include [IPS09; DO10]. Indeed, numerous multi-party computation implementations are available online, such as VIFF (`viff.dk`) [DGKN09].

In the common "pre-processing model", where pre-processing time is available before the main computation, yet faster protocols are possible: [DPSZ12] achieves secure 3-party 64-bit multiplication in 0.05 ms. This could be a very reasonable model when the same N players play multiple or repeated games.

We note that there has been a line of work starting with [DHR00], on designing multi-party computation protocols specifically for sampling from correlated equilibrium distributions. However, these address the two-party setting, and have not taken into account the most recent advances in general multi-party computation techniques, so we do not consider them to be of great relevance here.

Perfect setting. In the perfect setting, known protocols are less efficient; and perfectly secure encryption is relatively inefficient due to inherently large key sizes. Nonetheless, substantial progress has been made: the best known protocol [BTH08] achieves $O(N)$ communication complexity per multiplication[8], improving on previous protocols by $\Omega(N^2)$.

We consider our information-theoretic results to be of interest primarily as proofs of possibility, and a novel application of cryptographic techniques to game theory without computational restrictions. Certainly, for efficiency in practice and strength of results, our computational protocols are the ones of interest.

[8]The circuit that the parties want to compute is usually represented as addition and multiplication gates, and the multiplication gates have been found to be the bottleneck for multi-party computation.

6 Conclusion

In this work we use standard cryptographic tools – namely, encryption schemes – to introduce the concept of blinded games: strategic games in which players take encrypted actions, and in particular have the possibility to take actions they know nothing about. Moreover, we provide cryptographic protocols that enable the players to not rely on trusted mediators in order to achieve equilibrium payoffs.

Our approach suggest new interesting uses of cryptographic methods in game theory. We show that our blinded games offer a viable and appealing alternative to solution concepts based on commitment, and a particularly promising direction for future work is to apply the paradigm of leveraging players' lack of knowledge in order to avoid commitment, in broader settings.

Acknowledgements. We are grateful to Alessandra Scafuro for raising the question of encrypting advice, to Silvio Micali for very helpful advice on exposition, and to Jesper Buus Nielsen for detailed technical comments on the final versions.

Pavel Hubáček acknowledges support from the European Research Commission Starting Grant 279447; from the Danish National Research Foundation and The National Science Foundation of China (grant 61061130540) for the Sino-Danish Center for the Theory of Interactive Computation, within part of this work was performed; and from the CFEM research center, supported by the Danish Strategic Research Council.

Sunoo Park acknowledges support from NSF Eager CNS-1347364, NSF Frontier CNS-1413920, the Simons Foundation (agreement dated June 5, 2012), Air Force Laboratory FA8750-11-2-0225, and Lincoln Lab PO7000261954.

References

[Alw+09] Joël Alwen et al. "Collusion-Free Multiparty Computation in the Mediated Model". In: *Advances in Cryptology - CRYPTO 2009, 29th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2009. Proceedings*. Ed. by Shai Halevi. Vol. 5677. Lecture Notes in Computer Science. Springer, 2009, pp. 524–540. ISBN: 978-3-642-03355-1. DOI: 10.1007/978-3-642-03356-8_31. URL: http://dx.doi.org/10.1007/978-3-642-03356-8_31.

[AL11] Gilad Asharov and Yehuda Lindell. "A Full Proof of the BGW Protocol for Perfectly-Secure Multiparty Computation". In: *IACR Cryptology ePrint Archive* 2011 (2011), p. 136.

[Aum74] Robert J. Aumann. "Subjectivity and correlation in randomized strategies". In: *Journal of mathematical Economics* 1.1 (1974), pp. 67–96.

[AH03] Robert J. Aumann and Sergiu Hart. "Long cheap talk". In: *Econometrica* 71.6 (2003), pp. 1619–1660.

[Bár92] Imre Bárány. "Fair distribution protocols or how the players replace fortune". In: *Mathematics of Operations Research* 17.2 (1992), pp. 327–340.

[BTH08] Zuzana Beerliová-Trubíniová and Martin Hirt. "Perfectly-Secure MPC with Linear Communication Complexity". In: *Theory of Cryptography*. Ed. by Ran Canetti. Vol. 4948. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, pp. 213–230. ISBN: 978-3-540-78523-1. DOI: 10.1007/978-3-540-78524-8_13. URL: http://dx.doi.org/10.1007/978-3-540-78524-8_13.

[BDPR98] Mihir Bellare, Anand Desai, David Pointcheval, and Phillip Rogaway. "Relations Among Notions of Security for Public-Key Encryption Schemes". In: *CRYPTO*. Ed. by Hugo Krawczyk. Vol. 1462. Lecture Notes in Computer Science. Springer, 1998, pp. 26–45. ISBN: 3-540-64892-5.

[BGW88] Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. "Completeness theorems for non-cryptographic fault-tolerant distributed computation". In: *Proceedings of the twentieth annual ACM symposium on Theory of computing*. ACM. 1988, pp. 1–10.

[BP98] Elchanan Ben-Porath. "Correlation without Mediation: Expanding the Set of Equilibrium Outcomes by Cheap Pre-play Procedures". In: *Journal of Economic Theory* 80.1 (1998), pp. 108 –122. ISSN: 0022-0531. DOI: http://dx.doi.org/10.1006/jeth.1998.2397. URL: http://www.sciencedirect.com/science/article/pii/S0022053198923973.

[Bog+09] Peter Bogetoft et al. "Secure Multiparty Computation Goes Live". In: *Financial Cryptography and Data Security*. Ed. by Roger Dingledine and Philippe Golle. Vol. 5628. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, pp. 325–343. ISBN: 978-3-642-03548-7. DOI: 10.1007/978-3-642-03549-4_20. URL: http://dx.doi.org/10.1007/978-3-642-03549-4_20.

[CCD88] David Chaum, Claude Crépeau, and Ivan Damgard. "Multiparty Unconditionally Secure Protocols". In: *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*. STOC '88. Chicago, Illinois, USA: ACM, 1988, pp. 11–19. ISBN: 0-89791-264-0. DOI: 10.1145/62212.62214. URL: http://doi.acm.org/10.1145/62212.62214.

[Cle86] R Cleve. "Limits on the Security of Coin Flips when Half the Processors Are Faulty". In: *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*. STOC '86. Berkeley, California, USA: ACM, 1986, pp. 364–369. ISBN: 0-89791-193-8. DOI: 10.1145/12130.12168. URL: http://doi.acm.org/10.1145/12130.12168.

[DGKN09] Ivan Damgård, Martin Geisler, Mikkel Krøigaard, and Jesper Buus Nielsen. "Asynchronous Multiparty Computation: Theory and Implementation". In: *Public Key Cryptography*. Ed. by Stanislaw Jarecki and Gene Tsudik. Vol. 5443. Lecture Notes in Computer Science. Springer, 2009, pp. 160–179. ISBN: 978-3-642-00467-4.

[DO10] Ivan Damgård and Claudio Orlandi. "Multiparty Computation for Dishonest Majority: From Passive to Active Security at Low Cost". In: *CRYPTO*. Ed. by Tal Rabin. Vol. 6223. Lecture Notes in Computer Science. Springer, 2010, pp. 558–576. ISBN: 978-3-642-14622-0.

[DPSZ12] Ivan Damgård, Valerio Pastro, Nigel P. Smart, and Sarah Zakarias. "Multiparty Computation from Somewhat Homomorphic Encryption". In: *CRYPTO*. Ed. by Reihaneh Safavi-Naini and Ran Canetti. Vol. 7417. Lecture Notes in Computer Science. Springer, 2012, pp. 643–662. ISBN: 978-3-642-32008-8.

[DHR00] Yevgeniy Dodis, Shai Halevi, and Tal Rabin. "A Cryptographic Solution to a Game Theoretic Problem". In: *CRYPTO*. Ed. by Mihir Bellare. Vol. 1880. Lecture Notes in Computer Science. Springer, 2000, pp. 112–130. ISBN: 3-540-67907-3.

[DR07] Yevgeniy Dodis and Tal Rabin. "Cryptography and game theory". In: *Algorithmic Game Theory* (2007), pp. 181–207.

[DDN00] Danny Dolev, Cynthia Dwork, and Moni Naor. "Nonmalleable Cryptography". In: *SIAM J. Comput.* 30.2 (2000), pp. 391–437.

[FLP85] Michael J. Fischer, Nancy A. Lynch, and Michael S. Paterson. "Impossibility of Distributed Consensus with One Faulty Process". In: *J. ACM* 32.2 (Apr. 1985), pp. 374–382. ISSN: 0004-5411. DOI: 10.1145/3149.214121. URL: http://doi.acm.org/10.1145/3149.214121.

[FHM98] Matthias Fitzi, Martin Hirt, and Ueli M. Maurer. "Trading Correctness for Privacy in Unconditional Multi-Party Computation (Extended Abstract)". In: *Advances in Cryptology - CRYPTO '98, 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*. Ed. by Hugo Krawczyk. Vol. 1462. Lecture Notes in Computer Science. Springer, 1998, pp. 121–136. ISBN: 3-540-64892-5. DOI: 10.1007/BFb0055724. URL: http://dx.doi.org/10.1007/BFb0055724.

[GMW87] Oded Goldreich, Silvio Micali, and Avi Wigderson. "How to Play any Mental Game or A Completeness Theorem for Protocols with Honest Majority". In: *STOC*. Ed. by Alfred V. Aho. ACM, 1987, pp. 218–229. ISBN: 0-89791-221-7.

[GLR10] Ronen Gradwohl, Noam Livne, and Alon Rosen. "Sequential Rationality in Cryptographic Protocols". In: *FOCS*. IEEE Computer Society, 2010, pp. 623–632. ISBN: 978-0-7695-4244-7.

[HSHI02] Goichiro Hanaoka, Junji Shikata, Yumiko Hanaoka, and Hideki Imai. "Unconditionally Secure Anonymous Encryption and Group Authentication". English. In: *Advances in Cryptology ASIACRYPT 2002*. Ed. by Yuliang Zheng. Vol. 2501. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2002, pp. 81–99. ISBN: 978-3-540-00171-3. DOI: 10.1007/3-540-36178-2_5. URL: http://dx.doi.org/10.1007/3-540-36178-2_5.

[HNR13] Pavel Hubáček, Jesper Buus Nielsen, and Alon Rosen. "Limits on the Power of Cryptographic Cheap Talk". In: *CRYPTO (1)*. Ed. by Ran Canetti and Juan A. Garay. Vol. 8042. Lecture Notes in Computer Science. Springer, 2013, pp. 277–297. ISBN: 978-3-642-40040-7.

[IPS09] Yuval Ishai, Manoj Prabhakaran, and Amit Sahai. "Secure Arithmetic Computation with No Honest Majority". In: *Theory of Cryptography*. Ed. by Omer Reingold. Vol. 5444. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009, pp. 294–314. ISBN: 978-3-642-00456-8. DOI: 10.1007/978-3-642-00457-5_18. URL: http://dx.doi.org/10.1007/978-3-642-00457-5_18.

[ILM11] Sergei Izmalkov, Matt Lepinski, and Silvio Micali. "Perfect implementation". In: *Games and Economic Behavior* 71.1 (2011), pp. 121–140.

[KDG03] D. Kempe, A. Dobra, and J. Gehrke. "Gossip-based computation of aggregate information". In: *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*. 2003, pp. 482–491. DOI: 10.1109/SFCS.2003.1238221.

[Kra98] Hugo Krawczyk, ed. *Advances in Cryptology - CRYPTO '98, 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*. Vol. 1462. Lecture Notes in Computer Science. Springer, 1998. ISBN: 3-540-64892-5.

[LMPS04] Matt Lepinski, Silvio Micali, Chris Peikert, and Abhi Shelat. "Completely fair SFE and coalition-safe cheap talk". In: *PODC*. Ed. by Soma Chaudhuri and Shay Kutten. ACM, 2004, pp. 1–10. ISBN: 1-58113-802-4.

[LKC12] Joshua Letchford, Dmytro Korzhyk, and Vincent Conitzer. "On the value of commitment". In: *Autonomous Agents and Multi-Agent Systems* (2012), pp. 1–31.

[Mas99] Eric Maskin. "Nash Equilibrium and Welfare Optimality". In: *Review of Economic Studies* 66 (1999). Reprinted in J.J. Laffont (ed.), The Principal Agent Model: The Economic Theory of Incentives, London: Edward Elgar, 2003, pp. 23–38.

[MS09] Silvio Micali and Abhi Shelat. "Purely Rational Secret Sharing (Extended Abstract)". In: *TCC*. Ed. by Omer Reingold. Vol. 5444. Lecture Notes in Computer Science. Springer, 2009, pp. 54–71. ISBN: 978-3-642-00456-8.

[MRG13] Hervé Moulin, Indrajit Ray, and Sonali Sen Gupta. "Improving Nash by coarse correlation". In: *Journal of Economic Theory* 0 (2013), pp. –. ISSN: 0022-0531. DOI: http://dx.doi.org/10.1016/j.jet.2013.10.008. URL: http://www.sciencedirect.com/science/article/pii/S0022053113001798.

[MV78] Hervé Moulin and J-P Vial. "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon". In: *International Journal of Game Theory* 7.3-4 (1978), pp. 201–221.

[Nas50] John F. Nash. "Equilibrium points in n-person games". In: *Proceedings of the National Academy of Sciences* 36.1 (1950), pp. 48–49. DOI: 10.1073/pnas.36.1.48. eprint: http://www.pnas.org/content/36/1/48.full.pdf+html. URL: http://www.pnas.org/content/36/1/48.short.

[Sha79] Adi Shamir. "How to Share a Secret". In: *Commun. ACM* 22.11 (Nov. 1979), pp. 612–613. ISSN: 0001-0782. DOI: 10.1145/359168.359176. URL: http://doi.acm.org/10.1145/359168.359176.

[SZ10] Bernhard von Stengel and Shmuel Zamir. "Leadership games with convex strategy sets". In: *Games and Economic Behavior* 69.2 (2010), pp. 446–457.

[vS34] Heinrich von Stackelberg. *Marktform und Gleichgewicht.-Wien & Berlin: Springer 1934. VI, 138 S. 8.* J. Springer, 1934.

Appendix

A Extensive Games

Here we recall the standard definition of extensive games.

Definition A.1 (Extensive game)**.** *An extensive game* $\Gamma = \langle N, H, P, A, \mathcal{I}, (u_i) \rangle$ *is defined by:*

- *a finite set* N *of players,*
- *a set* H *of all possible history sequences (with the subset of all terminal histories denoted by* Z*),*
- *a player function* $P : H \setminus Z \to N$ *that assigns a player to every non-terminal history,*
- *a function* A *that assigns to every non-terminal history* $h \in H \setminus Z$ *a finite set of actions* $A(h) = \{a : (h, a) \in H\}$ *available to player* $P(h)$ *at* h*,*
- *for each player* $i \in N$*, a partition* $\mathcal{I}_i$ *of* $\{h \in H : P(h) = i\}$ *such that* $A(h) = A(h')$ *whenever* h *and* h' *are in the same* $I_i \in \mathcal{I}_i$*,*
- *for each player* $i \in N$*, a utility function* $u_i : Z \to \mathbb{R}$*,*

If the partition $\mathcal{I}_i$ is trivial and each $I_i \in \mathcal{I}_i$ contains a single history for every player i then we say that the extensive game is with *perfect information* (i.e., every player is perfectly informed of all actions taken by every other player). A strategy profile σ of an extensive game Γ with perfect information specifies the actions of every player at every history, i.e., for every $h \in H$ it specifies a probability distribution on $A(h)$ for player $i = P(h)$.

The solution concept relevant to this work in the context of extensive games with perfect information is Nash equilibrium.

Definition A.2 (Nash equilibrium of extensive game)**.** *Let $\Gamma = \langle N, H, P, A, (u_i)\rangle$ be an extensive game with perfect information. We say that strategy profile σ is a* Nash equilibrium of Γ *if for every player $i \in N$ and for every strategy σ_i' of player i:*

$$\mathbf{E}[u_i(\sigma)] \geq \mathbf{E}[u_i(\sigma_i', \sigma_{-i})],$$

where the expectations are taken over terminal histories sampled from the corresponding strategy profile.

Definition A.3 (Computational Nash equilibrium of extensive game)**.** *A* computational Nash equilibrium of extensive game $\Gamma = \langle N, H, P, A, (u_i)\rangle$ *is a* PPT*-samplable family of strategy profiles $\{\sigma^{(k)}\}_{k\in\mathbb{N}}$ for Γ if for every player $i \in N$ and for every* PPT*-samplable strategy σ_i' of player i it holds for all large enough k that*

$$\mathbf{E}[u_i(\sigma_i^{(k)})] \geq \mathbf{E}[u_i(\sigma_i', \sigma_{-i}^{(k)})] - \varepsilon(k),$$

where the expectations are taken over terminal histories sampled from the corresponding strategy profile, and ε is a negligible function.

In extensive games with *imperfect information*, the players are not informed about all the actions taken by their opponents. A *profile of behavioral strategies* specifies a probability distributions on actions available to every player $i \in N$ at every information set $I_i \in \mathcal{I}_i$. The solution concept of *Nash equilibrium in behavioral strategies* is defined similarly to Definition A.2.

When reasoning about Nash equilibrium in games with imperfect information we need to take into account also the beliefs of players about the past play at any information set. This gives rise to the notion of assessment.

Definition A.4 (Assessment)**.** *An* assessment *in an extensive game is a pair (β, μ), where β is a profile of behavioral strategies and μ is a funcion that assigns to every information set a probability distribution on the histories in the information set.*

The following solution concept aims to circumvent the instability of Nash equilibrium of extensive games with imperfect information.

Definition A.5 (Sequential equilibrium)**.** *Let $\Gamma = \langle N, H, P, A, \mathcal{I}, (u_i)\rangle$ be an extensive game. The assessment (β, μ) is a* sequential equilibrium *if it is*

1. sequentially rational - *for every $i \in N$, for every information set $I_i \in \mathcal{I}_i$, and every β_i'*

$$\mathbf{E}[u_i(\beta, \mu)|I_i] \geq \mathbf{E}[u_i((\beta_{-i}, \beta_i'), \mu)|I_i].$$

2. consistent - *there exists a sequence $\{(\beta^{(n)}, \mu^{(n)})\}_{n=1}^{\infty}$ of assessments that converges to (β, μ), $\beta^{(n)}$ is completely mixed for all $n \in \mathbb{N}$, and $\mu^{(n)}$ is derived from $\beta^{(n)}$ by Bayes' rule.*

B Verifiable Decryption

The following is the standard procedure for a *prover* to convince a *verifier* that he has correctly decrypted a ciphertext. Upon decrypting, the prover obtains the un-encrypted action and the randomness that was used during encryption, and presents the verifier with these two items. Then the verifier can run the encryption algorithm for herself, and check that the resulting ciphertexts are the same as the ones that they submitted for decryption. By the security of the encryption scheme, it would be (computationally) infeasible for the prover to come up with (decryption, randomness) pairs that pass this check, except by running the decryption algorithm with the correct secret key. Hence, the verifier may be assured that the prover has decrypted correctly.

Note that this verifiable decryption procedure requires that the encryption scheme, in addition to being secure[9], has the following property, which is very common among existing schemes:

- *Recoverable randomness.* By running the decryption algorithm on a ciphertext $c = \mathsf{Enc}(m)$ with a correct secret key, the decryptor must be able to recover the randomness used for encryption. More precisely, for any given keypair (pk, sk), we require that a decryptor possessing a correct secret key can efficiently compute *some* randomness r that, when inputted along with the correct message to the encryption algorithm, outputs the ciphertext in question, i.e. $\mathsf{Enc}_{pk}(m; r) = c$; and moreover, the decryptor cannot (with non-negligible probability) compute a randomness that, when inputted along with an incorrect message m' to the encryption algorithm, outputs the ciphertext in question, i.e. it is infeasible to find r' such that $\mathsf{Enc}_{pk}(m'; r') = c$.

We say that encryption schemes satisfying this property are *verifiably decryptable*.

C Implementation of *Any* CCE with Any Number of Players

Protocol 4. Implementing any computational correlated equilibrium α' of Γ^{Π}:

Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a CCA-secure public-key encryption scheme and let $(pk, sk) \leftarrow \mathsf{PGen}(1^k)$ with pk known to all players. Communication is via broadcast.

- The players run Protocol 1 as long as no player is detected to deviate from the protocol.
- If any player i is detected to deviate from the protocol, then all (other) players adopt joint min-max strategy (in Γ^{Π}) with the worst possible outcome for player i.

Theorem 5.4. Let $\Pi = (\mathsf{PGen}, \mathsf{PEnc}, \mathsf{PDec})$ be a CCA-secure public-key encryption scheme, and let Γ be any finite strategic game. For any coarse correlated equilibrium α of Γ, there exists a computational Nash equilibrium $\widetilde{\alpha}$ of the computational cheap talk extension $\widetilde{\Gamma^{\Pi}}$ that achieves the same utility profile as α.

Proof (sketch). The players run Protocol 4, in which – by construction – adhering to the protocol yields at least as much utility as deviation, for any given player. The full proof follows exactly the same structure as the proof of Theorem 1 of [DHR00], and we refer the reader to their paper for details. □

[9] CPA security suffices.

As discussed briefly in Section 5.2.1, implementation via Protocol 4 has the disadvantage (compared to Protocol 2) of *empty threats*: these are the subject of Appendix D.

D Protocol 2 is Free of Empty Threats

We start with an informal definition and discussion of empty threats, then proceed to a full definition and proof of empty-threat-freeness of Protocol 2.

Definition D.1 (Empty threat (informal))**.** *An empty threat posed by a player in an extensive game is a strategy of the threatening player at a history off the equilibrium path which is not rational from his perspective. A threatened player can demonstrate the existence of such an empty threat by taking a beneficial deviation that would make the threatening player refrain from following through with the announced threat.*

A consequence of empty threats in a Nash equilibrium is that a strategy profile containing empty threats is not sequentially stable, that is, players that adapt their strategies during the game would not follow such a strategy profile. One approach to avoid empty threats is to require subgame perfect equilibrium. However, as addressed by [GLR10], there is no obvious way to define subgame perfection in the computational setting. Therefore, we use the computational solution concept of [GLR10] and show that Protocol 2 is free of empty threats.

Now, we give the definition of empty-threat-free Nash equilibrium in extensive games (cf. [GLR10] for details and computational version). The definition uses the following notion of set of continuations of a strategy profile at a given history. For a history $h \in H$, a strategy σ, and a distribution $\tau = \tau(h)$ on $A(h)$, let

$$\mathrm{Cont}(h, \sigma, \tau) = \{\pi : (\pi \text{ differs from } \sigma \text{ only on the subgame } h) \& (\pi(h) = \tau(h))\}.$$

Definition D.2 (Empty threat)**.** *Let $\Gamma = \langle N, H, P, A, (u)_i \rangle$ be an extensive game, and let σ be a strategy profile. Then:*

- *For any history $h \in Z$, no player faces an empty threat at h with respect to σ.*
- *Player i* faces an empty threat at history $h \in H \setminus Z$ with respect to σ *if $i = P(h)$ and there exists a distribution $\tau = \tau(h)$ over $A(h)$ that satisfies the following: for all $\pi \in \mathrm{Cont}(h, \sigma, \tau)$ and $\pi' \in \mathrm{Cont}(h, \sigma, \sigma)$ for which no player faces an empty threat at any $h' \in H$ below h, it holds that*
$$\mathbf{E}[u_i(\pi)] > \mathbf{E}[u_i(\pi')].$$

A strategy profile σ is empty-threat-free on h *if for all $h' \neq \emptyset$ satisfying $(h, h') \in H$ no player faces an empty threat at (h, h') with respect to σ.*

Definition D.3 (Empty-threat-free Nash equilibrium)**.** *Let $\Gamma = \langle N, H, P, A, (u)_i \rangle$ be an extensive game. Strategy profile σ is an* empty-threat-free Nash equilibrium *if:*

- *σ is a Nash equilibrium of Γ, and*
- *for any $h \in H \setminus Z$, player $P(h)$ does not face an empty threat at h with respect to σ.*

Remark. The above definitions readily apply to extensive games with $n > 2$ players, even though they were originally intended for games with two players. As noted by [GLR10], a potential shortcoming of applying this definition in games with more than two players is that it does not take into account collusions between players. However, we accept that the definition addresses only unilateral deviations as is standard in Nash equilibrium.

Theorem D.4. *Let $\widetilde{\alpha}$ be the computational Nash equilibrium of $\widetilde{\Gamma^{\Pi}}$ from Theorem 5.3. Then $\widetilde{\alpha}$ is an empty-threat-free computational Nash equilibrium of $\widetilde{\Gamma^{\Pi}}$.*

Proof (sketch). Recall that the computational Nash equilibrium $\widetilde{\alpha}$ is payoff-equivalent to some computational coarse correlated equilibrium α of Γ (Theorem 5.3), and that α achieves at least as high utility for each player i as his worst Nash equilibrium σ^i of Γ to which the players default in case any deviation of player i is detected during the protocol phase.

For every security parameter k, we need to show that at any history no player is facing an empty threat (see Definition D.2), i.e. we need to show that there is no history h with a deviation τ such that every empty-threat-free continuation of τ improves over every empty-threat-free continuation of $\widetilde{\alpha}$ at h by more than negligible amount.

It follows from Lemma 4.6 that the expectation of any player after receiving the encrypted advice is the same as the expectation of playing $\widetilde{\alpha}$ without knowing the advice. Thus, the expectation from following the protocol is the same at any history of the cheap talk extension.

We will use the following claim that follows immediately from $\widetilde{\alpha}$ being a computational Nash equilibrium.

Claim D.5. *Any deviation during the protocol phase that goes unnoticed can give the player at most negligible advantage.*

Since any observed deviation corresponds to a history in which players default to the Nash equilibrium σ, no player is facing an empty threat with respect to $\widetilde{\alpha}$ at such histories, since σ is a Nash equilibrium.

By the definition of empty threat (Definition D.2), no player is facing an empty threat at the final round where players take simultaneous actions in the strategic game, and in particular it is an empty-threat-free strategy to play according to the received advice a_i' at the terminal history. By Claim D.5 any unobserved deviation in the protocol phase can yield at most negligible improvement in player's utility, thus we get by induction that to follow $\widetilde{\alpha}$ is an empty-threat-free continuation at any history.

Finally, Claim D.5 also gives that no continuation (in particular no empty-threat-free continuation) of any deviation at any history h can improve by more than negligible amount over the continuation induced at h by following $\widetilde{\alpha}$. Therefore, $\widetilde{\alpha}$ is an empty-threat-free computational Nash equilibrium of Γ^{Π}. □