Fairness with an Honest Minority and a Rational Majority

Shien Jin Ong* David Parkes† Alon Rosen‡ Salil Vadhan§

March 4, 2008

Abstract

We provide a simple protocol for secret reconstruction in any threshold secret sharing scheme, and prove that it is *fair* when executed with many *rational* parties together with a small minority of *honest* parties. That is, all parties will learn the secret with high probability when the honest parties follow the protocol and the rational parties act in their own self-interest (as captured by the notion of a Bayesian subgame perfect equilibrium). The protocol only requires a *standard* (synchronous) broadcast channel, and tolerates fail-stop deviations (i.e. early stopping, but not incorrectly computed messages).

Previous protocols for this problem in the cryptographic or economic models have either required an honest majority, used strong communication channels that enable simultaneous exchange of information, or settled for approximate notions of security/equilibria.

Keywords: game theory, fairness, secret sharing.

*Goldman, Sachs & Co., New York, NY. E-Mail: shienjin@alum.mit.edu. Work done while a graduate student at Harvard School of Engineering and Applied Sciences, supported by NSF grant CNS-0430336.

†Harvard School of Engineering and Applied Sciences, Cambridge, MA. E-Mail: parkes@eecs.harvard.edu

‡Herzliya IDC, Israel. E-Mail: alon.rosen@idc.ac.il. Work done in part while at Harvard's Center for Research on Computation and Society, School of Engineering and Applied Sciences.

§Harvard School of Engineering and Applied Sciences, Cambridge, MA. E-Mail: salil@eecs.harvard.edu. Work done in part while at UC Berkeley, supported by the Miller Institute for Basic Research in Science, a Guggenheim Fellowship, and NSF grant CNS-0430336.

1 Introduction

A major concern in the design of distributed protocols is the possibility that parties may deviate from the protocol. Historically, there have been two main paradigms for modeling this possibility. One is the cryptographic paradigm, where some parties are honest, meaning they will always follow the specified protocol, and others are malicious, meaning they can deviate arbitrarily from the protocol. The other is the economic paradigm, where all parties are considered to be rational, meaning that they will deviate from the protocol if and only if it is in their interest to do so.

Recently, some researchers have proposed studying mixtures of these traditional cryptographic and economic models, with various combinations of honest, malicious, and rational participants. One motivation for this that it may allow a more accurate modeling of the diversity of participants in real-life executions of protocols. Along these lines, the papers of Aiyer et al. [3], Lysyanskaya and Triandopoulos [25], and Abraham et al. [2] construct protocols that achieve the best of both worlds. Specifically, they take protocol properties that are known to be achievable in both the cryptographic model (with honest and malicious parties) and the economic model (with only rational parties), and show that protocols with the same properties can still be achieved in a more general model consisting of malicious and rational parties.

Our work is of the opposite flavor. We consider properties that are not achievable in either the cryptographic or economic models alone, and show that they can be achieved in a model consisting of both honest and rational parties. Specifically, we consider the task of secret reconstruction in *secret sharing*, and provide a protocol that is *fair*, meaning that all parties will receive the output, given rational participants together with a small minority of honest participants. In standard communication models, fairness is impossible in a purely economic model (with only rational participants) [19, 21] or in a purely cryptographic model (with a majority of malicious participants) [11]. Previous works in the individual models achieved fairness by assuming strong communication primitives that allow simultaneous exchange of information [19, 18, 2, 21, 23, 24, 20][1] or settled for approximate notions of security/equilibria [12, 8, 16, 30, 21], whereas we only use a standard (i.e. synchronous but not simultaneous) broadcast channel and achieve a standard notion of game-theoretic equilibrium (namely, a Bayesian subgame perfect equilibrium).

Thus, our work illustrates the potential power of a small number of honest parties to maintain equilibria in protocols. These parties follow the specified strategy even when it is not in their interest to do so, whether out of altruism or laziness. While we study a very specific problem (secret sharing reconstruction, as opposed to general secure function evaluation) in a simplified model (of only "fail-stop" deviations), we hope that eventually the understanding developed in this clean setting will be leveraged to handle more complex settings (as has been the case in the past).

Below, we review the cryptographic and economic paradigms in more detail. We then introduce the secret-sharing problem we study and survey recent works on this problem in the purely economic model. We then describe our results and compare them to what was achieved before.

1.1 The Cryptographic Paradigm

In the cryptographic paradigm, we allow for a subset of the parties to deviate from the protocol in an arbitrary, malicious manner (possibly restricted to computationally feasible strategies), and the actions of these parties are viewed as being controlled by a single adversary. Intuitively, this

[1] Actually, the impossibility results of [19, 21] also hold in the presence of a simultaneous broadcast channel and thus the works of [19, 18, 2, 21] use additional relaxations, such as allowing the number of rounds and/or the sizes of the shares to be unbounded random variables.

captures worst-case deviations from the protocol, so protocols protecting against such malicious and monolithic adversaries provide a very high level of security. Remarkably, this kind of security can be achieved for essentially every multiparty functionality, as shown by a series of beautiful results from the 1980's [35, 17, 9, 6, 31]. However, considering arbitrary (and coordinated) malicious behavior does have some important limitations. For example, it is necessary to either assume that a majority of the participants are honest (i.e. not controlled by the adversary) or allow for protocols that are unfair (i.e. the adversary can prevent some parties from getting the output). This follows from a classic result of Cleve [11], who first showed that there is no fair 2-party protocol for coin-tossing (even with computational security), and then deduced the general version by viewing a multiparty protocol an interaction between two super-parties, each of which controls half of the original parties. Lepinski et al. [23] bypass this impossibility result by assuming a strong communication primitive ("ideal envelopes") which allow simultaneous exchange of information, but it remains of interest to find ways of achieving fairness without changing the communication model.

1.2 The Economic Paradigm

In the economic paradigm, parties are modeled as rational agents with individual preferences, and will only deviate from the protocol if this is in their own self interest. This approach has become very popular in the computer science literature in recent years, with many beautiful results. There are two aspects of this approach:

1. Design computationally efficient mechanisms (i.e. functionalities that can be implemented by a trusted mediator) that give parties an incentive to be truthful about their private inputs, while optimizing some *social choice function*, which measures the benefit to society and/or the mechanism designer [27, 22, 5].

2. Implement these mechanisms by distributed protocols, with computational efficiency emphasized in *distributed algorithmic mechanism design* [13, 14, 15] and extended to also emphasize additional equilibrium considerations in *distributed implementation* [33, 28, 29], so that parties are "faithful" and choose to reveal private information as well as perform other message passing and computational tasks. More recent works achieve a strong form of distributed implementation, with provably no additional equilibria [24, 20], but require strong communication primitives and have less focus on computational efficiency.

Note that distributed algorithmic mechanism design is different in spirit from the traditional problem considered in cryptographic protocols, in that parties have "true" private inputs (whereas in cryptography all inputs are considered equally valid) and there is freedom to change how these inputs are mapped to outcomes (whereas in cryptography, the functionality is pre-specified.) Nevertheless, recent works have explored whether we can use the economic model to obtain 'better' solutions to traditionally cryptographic problems, namely to compute some pre-specified functionalities. One potential benefit is that we may be able to incentivize parties to provide their "true" private inputs along the lines of Item 1 above; the papers [26, 34] explore for what functionalities and kinds of utility functions this is possible.

A second potential benefit is that rational deviations may be easier to handle than malicious deviations (thus possibly leading to protocols with better properties), while also preferable to assuming a mixture of players at the honest and malicious extremes. This has led to a line of work, started by Halpern and Teague [19] and followed by [18, 2, 21], studying the problems of secret sharing and multiparty computation in the purely economic model, with all rational participants. One can also require notions of equilibria that are robust against coalitions of rational

players [2]. While this approach has proved to be quite fruitful, it too has limitations. Specifically, as pointed out in [18, 21], it seems difficult to construct rational protocols that are fair in the standard communication model, because parties may have an incentive to stop participating once they receive their own output. The works [19, 18, 2, 21], as well as [24, 20] applied to appropriately designed mediated games, achieve fairness by using strong communication primitives (simultaneous broadcast, "ballot boxes") that allow simultaneous exchange of information.

As mentioned above, we achieve fairness in the standard communication model by considering a mix of rational participants together with a *small* minority of honest participants. Note that Cleve's [11] proof that an honest majority is necessary in the cryptographic setting, by reduction to the two-party case, no longer applies. The reason is that we cannot view a subset of the rational parties as being controlled by a single super-party. Even in coalitional notions of equilibria, each individual in that subset would only agree to a coordinated (joint) deviation if it is in its own interest to do so.

Our protocol is for the share reconstruction problem in secret sharing, which we now describe in more detail.

1.3 Secret Sharing

In a *t-out-of-n secret-sharing scheme* [32, 7], a dealer takes a secret s and computes n (randomized) *shares* $s_1, \ldots, s_n$ of s, which are distributed among n parties. The required properties are that (a) any set of t parties can reconstruct the secret s from their shares, but (b) any set of fewer than t parties has no information about s (i.e. they would have been equally likely to receive the same shares for every possible value of s).

Secret sharing is a fundamental building block for cryptographic protocols [17, 6, 9, 31]. Typically, these protocols are structured as follows. First, every party shares its private input among all the parties. Then the computation of the functionality is done on shares (to maintain privacy). And at the end, the parties reveal their shares of the output so that everyone can reconstruct it. Our focus in this paper is on this final reconstruction step. Typically, it is assumed that there are enough honest parties in the protocol to ensure that the secret can be reconstructed from the revealed shares, even if some parties refuse to reveal their shares. A more challenging scenario is one where some parties may reveal incorrect values, which is handled by use of *verifiable secret sharing* [10], but for simplicity in this paper we only consider *fail-stop deviations*, where a party may stop participating in the protocol early but otherwise follows the prescribed strategy. (For example, this models people who may disconnect their computer from the network in the middle of the protocol, but do not have the time or skill to reprogram the software.) If we allow arbitrary fail-stop deviations, then it is clear that having $k \geq t$ honest parties are necessary and sufficient to have a reconstruction protocol that is *fair*, where everyone obtains the secret if anyone does. (In applications of secret sharing to secure multiparty computation, it is typically also important that the threshold is greater than the number of malicious parties, i.e. $t > n - k$. Combined with the previous statement, this implies that there are more honest parties than malicious ones, i.e. we need an honest majority.)

1.4 Rational Secret Sharing

It is natural to ask whether we can bypass this need for an honest majority by considering only *rational* deviations from the protocol. As noted above, the study of secret sharing with only rational participants was initiated by Halpern and Teague [19], and there have been several subsequent works [18, 21, 2]. In these works, it is assumed that participants prefer to learn the secret over

not learning the secret, and secondarily, prefer that as few other agents as possible learn it. As pointed out in Gordon and Katz [18], any protocol where rational participants reveal their shares sequentially will not yield a Nash equilibrium. This is because it is rational for the t'th player to stop participating, as she can already compute the secret from the shares of the first $t-1$ players and her own, and stopping may prevent the first $t-1$ players from learning it.

One way to get around this difficulty is to assume a *simultaneous broadcast channel*, where all parties can broadcast values at the same time, without the option of waiting to see what values the other parties are broadcasting. All parties simultaneously revealing their shares is a Nash equilibrium. That is, assuming all of the other parties are simultaneously revealing their shares, no party can increase her utility by aborting instead of revealing. This basic protocol is instructive because it has several deficiences:

1. A simultaneous broadcast channel is a strong (and perhaps unrealistic) communication primitive, particularly in the context of trying to achieve fairness, where the typical difficulties are due to asymmetries in the times that parties get information. For example, fair coin-tossing is trivial with a simultaneous broadcast channel (everyone broadcasts a bit, and the output is the exclusive-or), in contrast to Cleve's impossibility result for synchronous broadcast channels [11].

2. Nash Equilibrium in this context is a very weak guarantee. For example, as argued by Halpern and Teague [19], it seems likely that rational parties would actually abort. The reason is that aborting is never worse than revealing, and is sometimes better (if $t-1$ other parties reveal, then the tth party will always learn the secret and can prevent the other parties from doing so by an abort.)

Halpern and Teague [19] and follow-up works [18, 2, 21] focus on the second issue. That is, they allow simultaneous broadcast, and explore whether stronger solution concepts than plain Nash equilibrium can be achieved. Halpern and Teague [19] propose looking for an equilibrium that survives "iterated deletion of weakly dominated strategies." They prove that no bounded-round protocol can achieve a fair outcome in equilibrium when adopting this solution concept. However, they and subsequent works by Gordon and Katz [18] and Abraham et al. [2] show that fair outcomes are possible even with this equilibrium refinement using a probabilistic protocol whose number of rounds has finite expectation. Moreover, Abraham et al. [2] show how to achieve an equilibrium that is resistant to deviations by coalitions of limited size. Kol and Naor [21] argue that "strict equilibria" is a preferable solution concept to the iterated deletion notion used by Halpern and Teague [19], and show how to achieve it with a protocol where the size of shares dealt is an unbounded random variable with finite expectation. (They also show that a strict equilibrium cannot be achieved if the shares are of bounded size.) In all of the above works, the protocols' prescribed instructions crucially depend on the utilities of the various players.

The works of Lepinski et al. [24] and Izmalkov et al. [20] also can be used to obtain fair protocols for secret sharing by making an even stronger physical assumption than a simultaneous broadcast channel, namely "ballot boxes." Specifically, they show how to compile any game with a trusted mediator into a fair ballot-box protocol with the same incentive structure. Since the share-reconstruction problem has a simple fair solution with a trusted mediator (the mediator takes all the inputs, and broadcasts the secret iff *all* players reveal their share), we can apply their compiler to obtain a fair ballot-box protocol. But our interest in this paper is on retaining standard communication models.

1.5 Our Results

We focus primarily on the first issue: our goal is to achieve fairness without a simultaneous broadcast channel, but only a *synchronous broadcast channel*. That is, the protocol should proceed in rounds, and only one party can broadcast in each round.[2] When all parties are rational, the only positive result in this model is in work by Kol and Naor [21], who achieve a fair solution with an approximate notion of Nash equilibrium — no party can improve her utility by ε by deviating from the protocol. However, it is unclear whether such ε-Nash equilibria are satisfactory solution concepts (they seem "unstable"); indeed, Kol and Naor argue in favor of *strict* Nash equilibria, where parties will obtain strictly less utility by deviating (and show how to achieve strict equilibria in the presence of a simultaneous broadcast channel).

We instead assume that there is a *small* number k of honest participants (which can be much smaller than the secret-sharing threshold t), and the rest are rational. Our main result is that in this setting, there is a simple protocol that achieves fair outcomes under exact (and also strict) forms of Nash equilibria.

Our protocol is simple to describe. The participants take turns broadcasting their shares in sequence. However, if any of the first $t-1$ parties deviates from the protocol by stopping and refusing to broadcast her share, then the protocol instructs all subsequent parties to do the same. The intuition is that if there is likely to be at least one *honest* party after the first $t-1$ parties, then the first $t-1$ parties have an incentive to reveal their shares because if they do so, the honest party will also reveal her share and enable them to reconstruct the secret. Then we observe that if the set of honest parties is a random subset of $k = \omega(\log n)$ parties, then there will be an honest party after party $t-1$ with all but negligible probability, as long as $t \leq (1 - \Omega(1)) \cdot n$. Thus, assuming that parties have a nonnegligible preference to learn the secret, we obtain an *exact* equilibrium in which *everyone* learns the secret with all but negligible probability.

While this intuition is natural, it is somewhat delicate to model it game-theoretically. We introduce a framework of "extensive form games with public actions and private outputs," which enables us to model players' uncertainties about the inputs (i.e. shares) of other players as well as uncertainty about which players are honest and which are rational. (For simplicity, we assume that each player is honest independently with some probability p, but with small modifications, the result should extend to other distributions on the set of honest players.) This uncertainty is also incorporated into the solution concept we use, which is known as *Bayesian Nash equilibrium.* Actually, we achieve a *strict* Bayesian Nash equilibrium, as well as a Bayesian *subgame perfect* Nash equilibrium. The latter captures the idea that the strategy is rational to follow regardless of the previous history of messages; intuitively, this means that the equilibrium does not rely on irrational empty threats (where a player will punish another player for deviating even at his own expense). In addition, our protocol is significantly simpler than ones in earlier work [19, 18, 2, 21]. Finally, our protocol has a bounded number of rounds and does not require changing the underlying secret-sharing scheme.

As mentioned above, we only need a small minority of honest parties to achieve an equilibrium where everyone learns the secret with high probability. In contrast, if we had malicious participants rather than rational ones, then we would need $k \geq t$ honest parties for everyone to learn the secret. Following [21], we also show that the properties achieved by our protocol are impossible with only rational players. Thus, by considering a mixture of rational and honest players, we achieve something that is impossible in either the purely cryptographic or purely economic frameworks.

[2] For round efficiency, sometimes people use a slightly more general channel where many parties can broadcast in a single round, but deviating parties are can perform 'rushing' — wait to see what others have broadcast before broadcasting their own values.

2 Definitions

2.1 Games with Public Actions and Private Outputs

To cast protocol executions into a game-theoretic setting, we introduce the notion of *extensive games with public actions and private outputs*. The basis of this new notion is the more standard definition of *extensive form games with perfect information*. Extensive form games enable us to model the *sequential* nature of protocols, where each player considers his plan of action only following some of the other players' messages (the "*actions*" of the game-theoretic model). The perfect information property captures the fact that each player, when making any decision in the public phase of the protocol, is perfectly informed of all the events that have previously occurred. Thus, extensive form games with perfect information are a good model for communication on a *synchronous broadcast channel*.

We build upon extensive form games with perfect information and augment them with an additional final private stage. This additional stage models the fact that at the end of the game, each player is allowed to take some arbitrary action as a function of the history of messages so far. This action, along with the "history" of public actions that have taken place during the execution of the game is considered a "terminal" history (as well as the players' inputs), and has a direct effect on players' payoffs.

We work under the assumption that players $i \in N$ are handed private inputs θ_i that belong to some pre-specified set Θ_i (the "*types*" of the game-theoretic model) and specify a distribution according to which the various inputs are chosen (akin to *Bayesian* games). Players' inputs can be thought of as the shares for the secret-sharing scheme, and are generated jointly with the actual secret. The secret is thought of as a "reference" value that is not given to the players at the beginning of the protocol (but may be determined through messages exchanged), and is used at the output stage along with private actions to determine player utilities; in game-theoretic terms, this secret can be considered to be picked by "Nature" in the first round of the game and then induces a distribution μ on the private types of agents.

Definition 2.1 (Extensive game with public actions and private outputs)**.** *An* extensive form game with public actions and private outputs *is a tuple* $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ *where*

- $N = \{1, \ldots, n\}$ *is a finite set of* players,
- Δ *is a (possibly infinite) set of possible* reference values,
- $\Theta = \Pi_{i \in N} \Theta_i$, *where* Θ_i *is a (possibly infinite) set of possible* private types *of player* $i \in N$,
- *values* $(\delta, \theta_1, \ldots, \theta_n) \in \Delta \times \Theta$, *are chosen jointly according to the distribution* μ,
- H *is a (possibly infinite) set of (finite)* history *sequences satisfying that the empty word* $\epsilon \in H$. *The components of a history sequence* $h \in H$ *are called* public actions. *A history* $h \in H$ *is* terminal *if* $\{a : (h, a) \in H\} = \emptyset$. *The set of terminal histories is denoted* Z.
- $P : (H \setminus Z) \to N$ *is a function that assigns a "next" player to every non-terminal history.*
- A *is a function that assigns for every non-terminal history* $h \in H \setminus Z$ *and given a type* θ_i *of player* $i = P(h)$, *a finite set* $A(\theta_i, h) \subseteq \{a : (h, a) \in H\}$ *of available* public actions.
- F *is a (possibly infinite) set of* private actions *available to each player after a terminal history.*
- $u = (u_1, \ldots, u_n)$ *is a vector of payoff functions* $u_i : \Delta \times \Theta \times Z \times F^n \to \mathbb{R}$.

An extensive form game with public actions and private outputs is interpreted as follows: the reference value and the types of the players are selected according to the joint distribution μ. The type $\theta_i \in \Theta_i$ is handed to player $i \in N$ and the value δ is kept secret for future reference (it will affect the players' utilities). This is followed by a sequence of actions that are visible by all players. After any nonterminal history $h \in H \setminus Z$, player $P(h)$ chooses an action from the set $A(\theta_i, h)$. The empty history $h_0 = (\epsilon)$ is the starting point of the game. Player $P(\epsilon)$ chooses an action $a \in A(\theta_{P(\epsilon)}, \epsilon)$. This induces a history $h_1 = (a)$, and player $P(h_1)$ subsequently chooses an action from the set $A(\theta_{P(h_1)}, h_1)$; this choice determines the next player to move, and so on until a terminal history $h \in Z$ is reached. At this point, all players $i \in N$ simultaneously pick an action, b_i, from the set F, of available private actions The actions $A(\theta_i, h)$ for a nonterminal $h \in H \setminus Z$ are defined to depend on (private) type θ_i because we wish to model fail-stop deviations in which the only decision available to a player is whether to follow the suggested action of the protocol in determining which message to send or sending a "fail-stop" message. The utility or ("*payoff*") of player i for an execution of the game is then determined to be the value $u_i(\delta, \theta, h, b_1, \ldots, b_n)$.

2.2 Strategies

The action chosen by a player for every history after which it is her turn to move, is determined by her *strategy* function. As is required in extensive-form games, the strategy is defined for all histories, even ones that would not be reached if the strategy is followed. Given our notion of extensive form games with public actions and private outputs, we distinguish between the *public strategy* of a player and her *private strategy*. The former is applied to non-terminal histories and is what determines a player's actions during the execution of the public part of the game. The latter is applied to terminal histories and is what determines a player's output. Given $\theta_i \in \Theta_i$, a strategy for player $i \in N$ is thus a pair $s_i(\theta_i) = (m_i(\theta_i), f_i(\theta_i))$, where:

- The *public strategy* $m_i(\theta_i)$ is a function that takes a partial history $h \in H \setminus Z$ such that $P(h) = i$ and produces a public 'message', $m_i(\theta_i, h) \in A(\theta_i, h)$.
- The *private strategy* $f_i(\theta_i) : Z \to F$ takes a terminal history $h \in Z$ and produces a private 'output' $f_i(\theta_i, h) \in F$.

For notational simplicity we only discuss pure (i.e. deterministic) strategies; mixed (i.e. randomized) strategies can be modeled by adding additional 'coin tosses' to each player's type θ_i. This simplification is without loss of generality because our negative result holds for all secret-sharing reconstruction protocols (in particular, ones that also include additional coin tosses for mixed strategies), and we will show that our positive result (with a small number of honest players) has a pure strategy equilibrium.

We let $s = (s_1, \ldots, s_n)$ denote the vector of players' strategies, where $s_i = (m_i, f_i)$. Given a strategy vector s and the values $(\delta, \theta) = (\delta, \theta_1, \ldots, \theta_n)$ we define the outcome $o(\delta, \theta, s)$ of s given game Γ to be the tuple $(\delta, \theta, h, b_1, \ldots, b_n)$, where h is the terminal history $h \in Z$ that results when each player $i \in N$ is given a value $\theta_i \in \Theta_i$ that is sampled according to μ, publicly follows the actions chosen by m_i, and computes her final private output b_i using f_i. That is, h is a history $h = (a_1, \ldots, a_\ell)$ such that for $j = 1, \ldots, \ell - 1$ we have that $m_{P(a_1,\ldots,a_j)}(\theta_i, (a_1, \ldots, a_j)) = a_{j+1}$, and $b_i = f_i(\theta_i, h)$. The value of player i's utility is totally determined by the outcome $o(\delta, \theta, s)$. The initial distribution, μ, of the secret and the shares, along with the strategies $s_i = (m_i, f_i)$ induce a distribution on the outcome $o(\delta, \theta, s)$, and thus on the utilities of the players. Let $u_i(\mu, s)$ denote the expected value of player i's utility under distribution μ and strategy vectors s. We assume that rational players wish to maximize this value; this is formalized in the next section.

2.3 Bayesian Subgame Perfect Equilibrium

Let $s = (m, f)$ be a strategy vector in Γ. Define $u_i(\mu, s)$ to be the expected value of the utility of player $i \in N$, when the types are selected according to the distribution μ and all players follow strategy s. (Notice that for this to be well-defined in a game with histories of unbounded length, it must also hold that the game terminates with probability 1 when the players follow s.) Now, the notion of Nash equilibrium captures the idea that no player should have an incentive to change her strategy, assuming that the other players play s. That is, no change in player i's strategy can increase her expected utility. More formally, we require that $u_i(\mu, (s_{-i}, s'_i)) \leq u_i(\mu, s)$ for every s'_i, where (s_{-i}, s'_i) denotes the strategy vector where we replace the i'th component of s with s'_i.

Definition 2.2 (Nash equilibrium)**.** *A strategy profile s in the game $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ is said to be a* (Bayesian) Nash equilibrium *for Γ if*

1. *The game terminates with probability 1 when the player's types are selected according to μ and the players play according to s.*
2. *For all $i \in N$ and all strategies s'_i,*
   (a) *The game terminates with probability 1 when the players types are selected according to μ and the players play according to (s_{-i}, s'_i), and*
   (b) *$u_i(\mu, (s_{-i}, s'_i)) \leq u_i(\mu, s)$.*

We call s a strict Nash equilibrium *if the above holds when we instead require strict inequality in Condition (2b) (i.e. $u_i(\mu, (s_{-i}, s'_i)) < u_i(\mu, s)$) for every strategy s'_i that differs from s_i on a* public *action that occurs with nonzero probability when the players types are selected according to μ and the players play according to s.*

In game theory, the notion of Bayesian Nash equilibrium is typically formulated to say that for each $\theta_i \in \Theta_i$, player $i \in N$ maximizes her expected utility when the other players' types are distributed according to the conditional distribution $\mu|_{\theta_i}$. However, since a player's strategy depends on her type, this is equivalent to the formulation above, where we also take the expectation over θ_i.

The basic notion of Nash equilibrium (or even strict Nash equilibrium) turns out to be an unsatisfactory solution concept for extensive-form games. The reason is that a Nash equilibrium can rely on "incredible" threats by players — ones that are needed to maintain the equilibrium but never occur during the equilibrium play and would not be in the self-interest of the player if tested. A more appealing solution concept is that of *subgame perfect equilibrium*. This is a standard strengthening of the notion of Nash equilibrium in that it requires that the equilibrium strategy is a Nash equilibrium in every *subgame* of the original extensive game. We also take the standard approach of adapting the subgame perfect equilibrium concept to the Bayesian setting, and take into consideration the inherent uncertainty about player's types (and as a result about the actions implied by their strategies).

The assumption underlying the Bayesian setting is that individual players have *beliefs* about the values of other players' types. The beliefs are in fact distributions from which players think that the types of other players were drawn. At the beginning of the game, the belief corresponds to the initial distribution μ conditioned on the player's knowledge of her own type. As the game progresses, players update their beliefs as a function of other players' actions (recall that a player's actions may depend on her type). We make the following standard assumptions on the way in which the beliefs of players are updated:

- If a player takes an action that is consistent with equilibrium play, then we assume the player has followed the equilibrium strategy (since it is in her interest to do so) and other players update their beliefs by conditioning their previous beliefs on the action taken.
- If a player takes an action that is inconsistent with equilibrium play (i.e. would occur with probability 0 in equilibrium), then other players do not update their beliefs at all.

The notion of Bayesian subgame perfect equilibrium requires that the game remains in Nash equilibrium even after beliefs are updated as above following any sequence of actions. It seems that any reasonable solution concept would satisfy this requirement. If anything, one might want a stronger guarantee after witnessing non-equilibrium play, such as following the equilibrium strategy being in a player's interest regardless of the type of the player who deviated. (See [**?**] for further discussion of the difficulties related to Bayesian updating in extensive form games.)

The solution concept is a natural extension of the notion of subgame perfect equilibrium, adapted to the Bayesian setting. It builds on the following definition of a subgame, which recursively captures the way in which players update their beliefs as a result of other players' actions.

Definition 2.3 (Subgame)**.** *Let $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ be an extensive form game with public actions and private outputs and $s = (m, f)$ a strategy profile for Γ. For a history $h \in H$ the* subgame *$\Gamma(s, h) = (N, H|_h, P|_h, A|_h, F, \Delta, \Theta, \mu|_h, u|_h)$ and* substrategy *$s|_h$ are defined recursively as follows.*

1. *For the empty history $h = \epsilon$, we set $\Gamma(s, h) = \Gamma$ and $s|_h = s$.*
2. *For a history $h = a$ consisting of a single action a, we define $\Gamma(s, a)$ and $s|_a$ as follows:*
   - *$H|_a$ is the set of sequences h' for which: $(a, h') \in H$; i.e., $H|_a = \{h : (a, h') \in H\}$.*
   - *The function $P|_a$ is defined by $P|_a(h') = P(a, h')$ for each $h' \in H_a$.*
   - *The function $A|_a$ is defined by $A|_a(\theta_i, h') = A(\theta_i, (a, h'))$.*
   - *For every $i \in N$, the function $u_i|_a$ is defined by $u_i|_a(\delta, \theta, h', b) = u_i(\delta, \theta, (a, h'), b)$.*
   - *The distribution $\mu|_a$ is defined as follows: if a has nonzero probability in Γ when players follow strategy s and the types are chosen according to μ, then $\mu|_a$ is μ conditioned on a. Otherwise, $\mu|_a = \mu$.*
   - *The substrategy $s|_a = (m|_a, f|_a)$ is defined by $m_i|_a(\theta_i, h') = m_i(\theta_i, (a, h'))$ and $f_i|_a(\theta_i, h') = f_i(\theta_i, (a, h'))$.*
3. *For a history $h = (a, h')$ of length greater than 1 starting with an action a, we recursively define $\Gamma(s, h) = (\Gamma(s, a))(s|_a, h')$.*

Definition 2.4 (Bayesian Subgame Perfect Equilibrium)**.** *Let $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ be an extensive form game with public actions and private outputs. A strategy profile $s = (m, f)$ is said to be a* Bayesian subgame perfect equilibrium *for the game Γ if for every history $h \in H$, $s|_h$ is Nash equilibrium for the subgame $\Gamma(s, h)$.*

3 Secret-Sharing

In this section we formally define the idea of a secret sharing scheme. We then define what is a *secret sharing reconstruction protocol*, and cast it into a game-theoretic setting by defining a corresponding *fail-stop game* (as induced by a given reconstruction protocol). The latter allows us to reason about whether players have an incentive to follow the strategy specified by the protocol, assuming their only way of deviating is to fail (and change their private output).

3.1 Secret-Sharing Schemes

In order to define a secret sharing scheme we need to specify the sets from which the secret and its shares are drawn, as well as a (joint) probability distribution under which the shares of a secret are generated by the dealer (along with the secret).

Definition 3.1 (Secret sharing scheme)**.** *A* threshold secret sharing scheme *is a tuple* $(N, t, \Delta, \Theta, \mu, g)$ *where*

- $N = \{1, 2, \ldots, n\}$ *is the set of* players*,*
- $t \in \mathbb{N}$ *is the* threshold *of the scheme,*
- Δ *is the set from which the "secret" is chosen,*
- $\Theta = \Pi_{i \in N} \Theta_i$*, where* Θ_i *is the set of possible shares of player* $i \in N$*,*
- μ *is a joint probability distribution for the secret and the shares* $(\delta, \theta_1, \ldots, \theta_n) \in \Delta \times \Theta$*,*
- $g = \{g_S : \prod_{i \in S} \Theta_i \to \Delta\}_S$ *is a collection of* reconstruction functions*, such that for each set* $S \subseteq N$ *of size at least* t*:*
$$\Pr[g_S((\theta_i)_{i \in S}) = \delta] = 1$$
where $(\delta, \theta_1, ..., \theta_n)$ *are drawn according to* μ*, and*
- *for every* $S \subseteq N$ *of size less than* t*, the tuple* $(\delta, (\theta_i)_{i \in S})$ *has the same distribution as* $(U, (\theta_i)_{i \in S})$ *when* δ *and the* θ_i*'s are chosen according to* μ *and* U *is uniformly and independently chosen in* Δ*.*

A secret sharing scheme is implemented by letting a trusted *dealer* jointly pick the secret and shares according to the distribution μ, and then distributing share $\theta_i \in \Theta_i$ to player $i \in N$. The reconstruction functions are what enables any set S of at least t players to use their shares $(\theta_i)_{i \in S}$ in order to jointly reconstruct the secret. The scheme should also guarantee secrecy against any subset S of less than t players. This requirement is expressed in the last item of Definition 3.1.

3.2 Reconstruction Protocols

Once shares are distributed among the players, it is required to specify a protocol according to which the players can jointly reconstruct the secret at a later stage (using the reconstruction function). The reconstruction protocol prescribes a way in which the players compute their "messages", which are chosen from a given fixed "alphabet," and are then broadcasted to all other players. The protocol also specifies an *output function* that is used by the players to compute their (private) output. One specific way of doing so would be to let the players broadcast their private shares in some order. However, to avoid restrictions on the way in which the protocol proceeds, we give a more general definition that allows the exchange of arbitrary messages.

Definition 3.2 (Secret sharing reconstruction protocol)**.** *A* reconstruction protocol *for a secret sharing scheme* $(N, t, \Delta, \Theta, \mu, g)$ *is a tuple* $\Pi = (\Sigma, H, P, m^*, f^*)$ *where*

- Σ *is a finite set of* messages *that contains a special "deviation" symbol, which we denote by* $\perp$*.*
- $H \subseteq \Sigma^*$ *is a (possibly infinite) set of* protocol history *sequences satisfying that the empty word* $\epsilon \in H$*. We let* $M(h) = \{m : (h, m) \in H\} \subseteq \Sigma$*. A history* $h \in H$ *is* terminal *if* $M(h) = \emptyset$*. The set of terminal histories is denoted* Z*. We require that* $\perp \in M(h)$ *for all* $h \in H \setminus Z$*.*

- $P : (H \setminus Z) \to N$ *is a function that assigns a "next" player to every non-terminal history.*
- $m^* = (m_1^*, \ldots, m_n^*)$ *is a vector of* next-message functions *where for every* $\theta_i \in \Theta_i$, $m_i^*(\theta_i)$ *is a function mapping every history* $h \in H \setminus Z$ *such that* $P(h) = i$ *to a message* $m_i^*(\theta_i, h) \in M(h) \setminus \{\perp\}$, *unless* h *has a prefix* $(h', \perp)$ *such that* $P(h') = i$ *(i.e. player* i *has already deviated). In the latter case, we require that* $m_i^*(\theta_i) = \perp$.
- $f^* = (f_1^*, \ldots, f_n^*)$ *is a vector of* output functions $f_i^* : \Theta_i \times Z \to \Delta$.

Some comments about the choices made in the definition are in place:

- As we discussed earlier in the context of pure vs. mixed strategies, we assume without loss of generality that the player's next message function is deterministic. This is because any required randomness for computing the next message function could have been incorporated into the choice of the share θ_i (that is distributed by the dealer). Similarly, we may also incorporate in this way any auxiliary randomness that may be required in the protocol. This randomness can be thought of as being drawn jointly by the dealer (or Nature for that matter) independently of the shares and secretly handed to the players along with the shares.
- The special "deviation" symbol, $\perp$, is introduced to allow "fail-stop" deviations from the protocol — if a party sends $\perp$, this means that she has stopped participating in the protocol (e.g. not sent a message within a prescribed amount of time). The protocol never asks a party to stop (since $\perp$ is required to be outside the range of m_i), and thus all parties can recognize when a party has deviated in this manner. However, the protocol still continues after such a deviation, as captured by the fact that the next-message functions are well-defined even on histories that contain $\perp$.
- Note that once a player has sent $\perp$, we do *not* banish her from the protocol. She continues to hear subsequent broadcasts on the channel, though she may not broadcast any messages other than $\perp$ in the future. We note that our negative results (Theorem 4.3) actually hold even if banishing is allowed, but our positive result (Theorem 5.2) does not require banishing.

A reconstruction protocol for a given secret sharing scheme is implemented under the assumption that the secret and shares $(\delta, \theta_1, \ldots, \theta_n)$ are chosen according to the distribution μ. The protocol is interpreted as follows: after any non-terminal protocol history $h \in H \setminus Z$, player $i = P(h)$ chooses a message $m = m_i^*(\theta_i, h) \in \Sigma \setminus \{\perp\}$. The empty history $h_0 = \epsilon$ is the starting point of the game. Player $i_0 = P(\epsilon)$ chooses a message $m = m_{i_0}^*(\theta_{i_0}, \epsilon) \in \Sigma \setminus \{\perp\}$. This induces a history $h_1 = (m)$, and player $P(h_1)$ subsequently chooses a message from the set $\Sigma \setminus \{\perp\}$; this choice determines the next player to move, and so on until a terminal history $h \in Z$ is reached. At this point all players can determine the value of their private output functions, $f_i^*(\theta_i, h)$. Generally, we are interested in secret-sharing protocols in which all players will compute the secret correctly (i.e. $f_i^*(\theta_i, h) = \delta$ with high probability over μ, provided all players follow the protocol). Rather than require this as part of the definition, however, we will address explicitly this in the statements of our positive and negative results.

3.3 Fail-Stop Games

Based on the definition of a secret sharing protocol, we may now formalize an induced *fail-stop game*. Loosely speaking, this is an interpretation of a reconstruction protocol as an extensive form game with public messages and private outputs, in which only *fail-stop deviations* are allowed: a

player may choose not to follow the protocol's instructions, but her choice is limited to *whether to stop or not*, and in case she decides to stop, the deviation is *visible by all other players*.

Generally speaking, the definition of a fail-stop game consists of a natural interpretation of the protocol as a game: protocol histories correspond to histories of an extensive form game, messages in the protocol correspond to actions in the game, next message functions correspond to strategies, and the outputs correspond to output actions.

We model a situation in which the only choice a player has in the fail-stop game is whether to continue with the prescribed instructions (and in particular choose an action according to m_i^*), or to deviate from Π (by sending the $\perp$ message). Thus, at each point of the game, the action space available to a player is the subset Σ that consists of "legitimate" protocol messages along with the $\perp$ message. Having a single possible message allocated for "deviation" is precisely what captures the fact that a player is limited to a single type of deviation from the protocol's instructions.

Definition 3.3 (Fail-stop game for secret sharing)**.** *A* fail-stop game *that corresponds to the reconstruction protocol $\Pi = (\Sigma, H, P, m^*, f^*)$ for a secret sharing scheme $(N, t, \Delta, \Theta, \mu, g)$ is an extensive form game with public actions and private outputs $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ satisfying the following conditions:*

- *The set of private actions available to each player is $F = \Delta$.*
- *For every nonterminal history $h \in H \setminus Z$ and every $\theta_i \in \Theta_i$, the set of available* public actions *to player $i = P(h)$, is $A(\theta_i, h) = \{m_i^*(\theta_i, h), \perp\}$.*
- *For an outcome $o = (\delta, \theta, h, b_1, \ldots, b_n)$, the utilities $u_i(o)$ are a function of only i and the set $S(o) = \{j : b_j = \delta\}$. Moreover, we require that:*
  1. *If $i \in S(o)$ and $i \notin S(o')$, then $u_i(o) > u_i(o')$,*
  2. *If $S(o) \subsetneq S(o')$ and either $i \in S(o) \cap S(o')$ or $i \notin S(o) \cup S(o')$, then $u_i(o) > u_i(o')$.*

The honest strategy vector *in Γ is the pair $s^* = (m^*, f^*)$.*

A fail stop game is interpreted as follows: a dealer selects a secret and shares from the joint distribution μ and hands the share $\theta_i \in \Theta_i$ to player $i \in N$. This is followed by a sequence of actions (messages) that are prescribed by the reconstruction protocol Π, and are in particular visible by all players. At each point $h \in H$ of the game the player $i = P(h)$, whose turn to play is next faces a decision of whether to continue according to the prescribed strategy m_i^*, or to deviate from the prescribed instruction. In case the player has chosen to follow the strategy, she broadcasts the message $m_i^*(\theta_i, h)$. Otherwise, she is considered to have chosen the action $\perp$. In both cases the history h is updated accordingly. Note that a deviation action $\perp$ is included in the history. Thus, all players can determine whether a certain player has deviated from the protocol. Once a terminal history h is reached, we allow each player to choose any private output from $F = \Delta$. That is, while we restrict public actions to be either according to the honest strategy m_i^* or failure ($\perp$), we place no restrictions on the private output. This is analogous to the "semi-honest" (aka "honest-but-curious") adversary model often studied in cryptography.

4 Impossibility Results

In this section, we prove impossibility results showing that fail-stop games for secret sharing have no equilibria satisfying certain properties. Later we will show that these properties are achievable

if we assume that a small number of the participants are *honest* (rather than rational). One of the properties we require is that the equilibrium satisfies one of the solution concepts from the previous section — either being a strict Nash equilibrium or being a Bayesian subgame perfect equilibrium. One additional property we will require is that in equilibrium, parties learn the secret with certainty, where this event occurs with nonzero probability. Formally,

Definition 4.1 (Knowing the secret with certainty)**.** *Let $s = (m, f)$ be a strategy profile for a fail-stop game $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$, let $h \in Z$ be a terminal history and $\theta_i \in \Theta_i$ be a type for player i. We say that player i* knows the secret with certainty at *(h, θ_i) if it holds that $b_i = \delta$ for every outcome $o = (\delta, \theta, h, b_1, \ldots, b_n)$ containing (h, θ_i) that occurs with nonzero probability when the types are chosen according to μ and the players follow s. For an outcome $o = (\delta, \theta, h, b_1, \ldots, b_n)$, we say that player i* knows the secret with certainty at *o if player i knows the secret with certainty at the pair (h, θ_i) contained in o.*

This definition is fairly natural in the context of fail-stop games. If the player's messages in the protocol consist of entire shares under the secret-sharing scheme (possibly with some 'coordination' information that does not depend on the shares), then a player will have learned the secret with certainty if she has seen at least $t - 1$ shares other than her own. However, it may not hold in protocols where player's reveal their shares gradually, e.g. the 'gradual release' protocols of [12].

The next condition captures the idea that 'stop' symbol, $\perp$, should not be used to convey information about a player's share or type. This is consistent with the 'fail-stop' spirit, and in particular will hold for the honest strategy profile.

Definition 4.2 (Fail-stop admissibility)**.** *A strategy profile $s = (m, f)$ for a fail-stop game $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ is* fail-stop-admissible *if for every player $i \in N$, non-terminal history $h \in H \backslash Z$ such that $P(h) = i$, it holds that if $m_i(\theta_i, h) = \perp$ for some type $\theta_i \in \Theta_i$, then $m_i(\theta_i', h) = \perp$ for all $\theta_i' \in \Theta_i$.*

Theorem 4.3. *Let $\Pi = (\Sigma, H, P, m, f)$ be a reconstruction protocol for a secret-sharing scheme $(N, t, \Delta, \Theta, \mu, g)$, with $1 < t < |N|$ and $|\Delta| > 1$, and let $\Gamma = (N, H, P, A, F, \Delta, \Theta, \mu, u)$ be a fail-stop game corresponding to Π. Then:*

1. *In every strict Nash equilibrium s of Γ, it is always the case that someone does not know the secret with certainty. That is for every outcome o that occurs with nonzero probability when the players' types are chosen according to μ and the players follow s, there is a player i who does not know the secret with certainty at o.*

2. *In every Bayesian subgame perfect equilibrium s of Γ that is fail-stop-admissible, it is always the case that someone does not know the secret with certainty.*

The proofs of our impossibility results are similar to, and indeed inspired by, those of Kol and Naor [21], but do not appear to follow from the statements of their results. Their impossibility results apply to protocols in which all players always compute the secret correctly (i.e. with probability 1), while we relax this to only having a nonzero probability that the players learn the secret with certainty. Their results also either require that the shares are taken from a finite domain (while ruling out protocols even with a simultaneous broadcast channel) or are restricted to two players (for shares from a countable domain and a synchronous broadcast channel), while our results have no constraint on the share domain and apply for any number of players (for a synchronous broadcast channel).

Proof. The proofs of both parts rely on the following notion. For every history $h \in H$, we define K_h to be the set of players that know the secret (with certainty) at h. Formally, we consider the subgame $\Gamma(s,h) = (N, H|_h, P|_h, A|_h, F, \Delta, \Theta, \mu|_h, u|_h)$ and say that player i *knows* the secret at h if there is a function f_i such that $f_i(\theta_i) = \delta$ with probability 1 over $(\delta, \theta_1, \ldots, \theta_n)$ chosen according to $\mu|_h$. K_h is the set of players that know the secret at h. Observe that this formulation is consistent with Definition 4.1 if s is a Nash equilibrium strategy, since players prefer to compute the secret if they can.

Observe that:

- Initially, no player knows the secret. That is, $K_\epsilon = \emptyset$. This follows from the secrecy property of the secret-sharing scheme and the fact that $t > 1$.
- With each additional play, the set of players that know the secret can only increase. That is, for every $(h,a) \in H$, we have $K_h \subseteq K_{(h,a)}$. This is because $\mu|_{(h,a)}$ is obtained by conditioning $\mu|_h$ (or equals $\mu|_h$ in case a is inconsistent with equilibrium play).

For both parts of the theorem, we begin by assuming for contradiction that there exists an outcome o^* (that occurs with nonzero probability under μ and s) where every player knows the secret with certainty. Thus $K_{h^*} = N$ for the corresponding terminal history h^*. Now we proceed to consider the two parts of the theorem separately.

1. Since $K_\epsilon = \emptyset$, there must exist a prefix (h', a) of h^* such that $K_{h'} \neq N$ and $K_{(h',a)} = N$. Let $i = P(h')$ be the player whose turn it is to move at h'. We observe that player i cannot have already stopped (i.e. sent $\perp$) in h'. Otherwise, $a = \perp$ would be the only action available to i at h' (regardless of the value of θ_i), and thus we would have $\mu|_{(h',a)} = \mu_{h'}$ and $K_{(h',a)} = K_{h'}$. Since i cannot have already stopped in h', it means that i has an action $a' \neq a$ available to it at h'. But the utility that i can obtain by playing a' is at least as good as playing a. In both cases, player i will compute the secret correctly with certainty, and if i plays a', then all other players will compute the secret correctly with certainty (in any optimal strategy). This contradicts the strictness of the equilibrium.

2. For this part, we prove the following statement for every subgame $\Gamma(s,g)$ of Γ, by induction on the number of players that have not previously stopped (sent $\perp$) in g.

   Claim 4.4. *If $K_g \neq N$, then $K_{(g,h)} \neq N$ for every terminal history h that occurs with nonzero probability in $\Gamma(s,g)$ when the types are chosen according to $\mu|_g$ and players play according to $s|_g$.*

   Notice that applying the claim with $g = \epsilon$ completes the proof of the theorem.

   > *Proof of Claim:* First, we consider the base case, when all players have already stopped in the history g. Then all players must always send $\perp$ in $\Gamma(s,g)$, and thus conditioning on a further transcript h does not change who knows the secret. That is, $\mu|_{g,h} = \mu|_g$ and $K_{g,h} = K_g \neq N$ for every history h that occurs with nonzero probability in $\Gamma(s,g)$ when the types are chosen according to $\mu|_g$ and players play according to $s|_g$.
   >
   > For the inductive step, consider a subgame $\Gamma(s,g)$ that violates the claim for sake of contradiction. Then, as in Part 1, there is a prefix (h', a) of h such that $K_{(g,h')} \neq N$ but $K_{(g,h',a)} = N$. Let $i = P|_g(h') = P(g,h')$ be the player whose

turn it is after h'. Now, we observe that $\perp$ must occur with zero probability as the message after h' when the players' types are drawn according to $\mu|_g$ and the players play according to $s|_g$ in $\Gamma(s,g)$: otherwise, by fail-stop-admissibility, player i must always play $\perp$ after h' (regardless of her type) and we would have $K_{(g,h',a)} = K_{g,h'}$.

Since $\perp$ is outside the support of equilibrium play, we have $\mu|_{(g,h',\perp)} = \mu|_{(g,h')}$ and $K_{(g,h',\perp)} = K_{(g,h')} \neq N$. Moreover, one more player has stopped in $\Gamma(s,(g,h'))$ than in $\Gamma(s,g)$ (namely player i). By induction $K_{(g,h)} \neq N$ for every terminal history h occurring with nonzero probability in $\Gamma(s,(g,h))$. This implies that player i's utility in $\Gamma(s,g)$ will be strictly higher if she plays $\perp$ instead of a after h', violating the fact that s is a Bayesian subgame perfect equilibrium. ■

■

5 Our Protocol

Our goal is to show that every secret-sharing scheme has a reconstruction protocol so that any fail-stop game that corresponds to it has an equilibrium strategy in which all players learn the secret. However, by the impossibility result of Section 4 this cannot be achieved if all players are rational (at least if the equilibrium is fail-stop admissible and the players know the secret with certainty).

5.1 Introducing an Honest Minority

To get around the impossibility result, we require that a small subset of *honest* players in the fail-stop game always follows the strategy prescribed by the reconstruction protocol (whether or not this is the best response to other players' actions). We model this scenario by assuming that the set of honest players is selected according to some distribution that will specify to each player whether she is to act honestly or rationally. The set of actions of an honest player will be then restricted to coincide with the strategy prescribed by the reconstruction protocol. The set of actions of a rational player will remain unchanged (i.e., as in the original definition of a fail-stop game).

Definition 5.1 (Fail-stop game with honest players)**.** *Let $\Pi = (\Sigma, H, P, m^*, f^*)$ be a reconstruction protocol for a secret sharing scheme $(N, t, \Delta, \Theta, \mu, g)$. A* fail-stop game that corresponds to Π with honest players *is an extensive form game with public actions and private outputs $\Gamma = (N, H, P, A, F, \Delta, \Theta \times \Omega, \mu \times \zeta, u)$, satisfying the following conditions:*

- *$\Omega = \Pi_{i \in N} \Omega_i$, where $\Omega_i = \{\texttt{honest}, \texttt{rational}\}$ indicates whether player $i \in N$ is honest or rational, and $\Theta_i \times \Omega_i$ is the set of possible types for player $i \in N$,*
- *ζ is a distribution on Ω, and values $(\delta, \theta, \omega) = (\delta, \theta_1, \ldots, \theta_n, \omega_1, \ldots, \omega_n) \in \Delta \times \Theta \times \Omega$ are chosen according to the distribution $\mu \times \zeta$. We refer to ζ as the* honest-player distribution.
- *The set of private actions available to each player is $F = \Delta$.*
- *For every nonterminal history $h \in H \setminus Z$ and every $\theta_i \in \Theta_i$, the set of available* public actions *to player $i = P(h)$, is*

$$A(\theta_i, \omega_i, h) = \begin{cases} \{m_i^*(\theta_i, h)\} & \text{if } w_i = \texttt{honest} \\ \{m_i^*(\theta_i, h), \perp\} & \text{if } w_i = \texttt{rational} \end{cases}$$

- *For an outcome $o = (\delta, (\theta, \omega), h, b_1, \ldots, b_n)$, the utilities $u_i(o)$ are a function of only i and the set $S(o) = \{j : b_j = \delta\}$. Moreover, we require that:*

  1. *If $i \in S(o)$ and $i \notin S(o')$, then $u_i(o) > u_i(o')$,*
  2. *If $S(o) \subsetneq S(o')$ and either $i \in S(o) \cap S(o')$ or $i \notin S(o) \cup S(o')$, then $u_i(o) > u_i(o')$.*

The honest strategy vector *in Γ is the pair $s^* = (m^*, f^*)$.*

We interpret a fail-stop game with honest players as follows. The private type of player $i \in N$ consists of a pair $(\theta_i, \omega_i) \in \Theta_i \times \Omega_i$ that is drawn along with other player's types and the reference value δ according to the distribution $\mu \times \zeta$. The value of $\omega_i \in \{\texttt{honest}, \texttt{rational}\}$ determines whether player i is bound to follow the honest strategy (as prescribed by Π), or will be allowed to deviate from it. The constraints on the set of actions of each player create a situation in which rational players are indeed free to deviate from the public strategy vector m^* (since they are allowed to choose the action $\perp$), whereas the honest players are in fact restricted to choose one of the available prescribed actions.[3] Note that we do not restrict the private actions of honest players; they may compute their output as an arbitrary function of the terminal history and their type. This is analogous to the notion of 'honest-but-curious' adversaries considered in the study of cryptographic protocols.

5.2 Our Main Result

We will show that the existence of a small number of honest players is sufficient to bypass the impossibility result proved in Theorem 4.3. That is, there is a reconstruction protocol such that every corresponding fail-stop game has an equilibrium in which such that with high probability all players learn the secret with certainty, provided that the set of honest players is uniform among all sets of a sufficiently large size and every player has a nonnegligible preference for learning the secret. The equilibrium is also fail-stop admissible, where the notion of fail-stop admissible is generalized to fail-stop games with honest players in a natural way: we require that sending $\perp$ does not reveal information about a player's share θ_i but allow it to reveal information about whether a player is honest or rational (indeed, only rational players can send $\perp$). (Formally, we require that if $m_i(\theta_i, \texttt{rational}, h) = \perp$ for some type $\theta_i \in \Theta_i$, then $m_i(\theta_i', \texttt{rational}, h) = \perp$ for all $\theta_i' \in \Theta_i$.)

Then our theorem is the following:

Theorem 5.2. *Every secret-sharing scheme $(N, t, \Delta, \Theta, \mu, g)$, with $t < |N|$, has a reconstruction protocol $\Pi = (\Sigma, H, P, m, f)$ such that the following holds. Let $\Gamma = (N, H, P, A, F, \Delta, \Theta \times \Omega, \mu \times \zeta_k, u)$ be a fail-stop game that corresponds to Π with honest players, where ζ_k is a distribution over tuples $(\omega_1, \ldots, \omega_n) \in \Omega$ for which $\omega_i = \texttt{honest}$ with probability k/n independently for all $i \in N$, for some real number $k \in [0, n]$. Suppose that for every $i \in N$:*

$$u_i(o_N) - u_i(o_\emptyset) > p(|N|, k) \cdot (u_i(o_N) - u_i(o_{N \setminus \{i\}})) + (1/|\Delta|) \cdot (u_i(o_{\{i\}}) - u_i(o_\emptyset)) \tag{1}$$

where o_S denotes an outcome where S is the set of players who compute the secret correctly and $p(n, k) = (1 - k/n)^{n-t+1} \leq \exp(-k \cdot (n - t)/n)$. Then Γ has a 'rational' strategy profile $s = (m, f)$ such that:

1. *s is fail-stop admissible,*

[3] Given θ_i and h the action $m_i^*(\theta_i, h)$ of player i is fully determined. Thus, in case that $\omega_i = \texttt{honest}$, player i has only one action to choose from (as determined by m_i^*), whereas in case that $\omega_i = \texttt{rational}$ player i has two actions to choose from (either the action determined by m_i^* or $\perp$).

2. *s is a strict Nash equilibrium in Γ,*

3. *s is a Bayesian subgame Nash equilibrium in Γ, and*

4. *The probability that all players learn the secret with certainty in Γ is at least $1 - p(|N|, k)$, when the players' types are chosen according to μ and they follow strategy vector s.*

We make a few remarks on the interpretation of this theorem:

- In the common case that $t \leq (1 - \Omega(1)) \cdot n$, observe that $p(n, k) = \exp(-\Omega(k))$ is negligible provided that $k = \omega(\log n)$, i.e. the expected number of honest players is superlogarithmic. In the common case that $|\Delta|$ is superpolynomial in n, Condition 1 simply says that a player's preference for learning the secret should not be negligible.

- Although the theorem allows the rational strategy profile $s = (m, f)$ to depend on the choice of the fail-stop game Γ (in particular its utility functions), it actually is mostly independent of Γ. The only dependence on the utility functions is in the private strategies f_i on histories that occur with probability at most $p(|N|, k)$.

- The honest-player distribution ζ_k models a situation in which players have no information about which other players are honest or rational, except for some a priori belief on the probability a given player is honest. With small modifications, the theorem should extend to other distributions ζ as well; see Remark 6.1.

5.3 The reconstruction protocol

Let $(N, t, \Delta, \Theta, \mu, g)$ be a secret-sharing scheme with $t < |N|$ and $|\Delta| > 1$. We assume that a dealer distributes shares to the players according to the distribution μ, and would like to design a protocol $\Pi = (\Sigma, H, P, m, f)$ for secret share reconstruction.

The protocol proceeds in a sequence of rounds, where in each round a single player can broadcast her share to all other parties (using a *synchronous broadcast* channel, as modelled by our definitions of reconstruction protocols (Def. 3.2) and extensive games with public actions and private outputs (Def. 2.1)). The order in which the players proceed is fixed in some arbitrary manner. For the sake of concreteness, suppose that at round i of the protocol, it is the turn of player i to broadcast. The protocol will instruct her to either *reveal* her share $\theta_i \in \Theta_i$ or *abort* the protocol by by broadcasting a special reserved symbol, which we denote by `ABORT`. Specifically, we will require that the next player i reveals θ_i unless one of the first $t - 1$ parties to speak (i.e. parties $1, \ldots, t - 1$) has either sent an `ABORT` message or deviated from the protocol's instructions (by broadcasting the special $\perp$ symbol). In the latter two cases player i should abort (by broadcasting the `ABORT` symbol).

After up to n such rounds, each player will locally use a reconstruction function $g_S \in g$ in order to try and compute the secret given the shares that have been revealed during the protocol's execution. By the properties of secret sharing, it follows that a party will be able to compute the secret (with certainty) at the end of the protocol if a set $S \subseteq N$ of at least $t - 1$ *other* parties have revealed their shares, and otherwise she has no information about the secret (i.e. can compute it with probability only $1/|\Delta|$). If t parties have revealed their shares, then we halt the public portion of the protocol, since all future moves are irrelevant (everyone will be able to compute the secret with certainty). Removing these irrelevant moves will enable us to argue that we achieve a strict Nash equilibrium.

In accordance with Definition 3.2, a player will be also allowed to deviate from the protocol's prescribed instructions by sending the special "fail-stop" $\perp$ symbol. (In an implementation, a player

that fails to broadcast her value within some predetermined amount of time might be considered to have broadcast the $\perp$ message.)

Protocol 5.3 (Reconstruction protocol)**.** *Given a secret sharing scheme $(N, t, \Delta, \Theta, \mu, g)$, we specify a reconstruction protocol $\Pi = (\Sigma, H, P, m, f)$ as follows:*

- $\Sigma = \left(\bigcup_{i \in N} \Theta_i\right) \cup \{\texttt{ABORT}, \perp\}$*, where* $\texttt{ABORT}, \perp \notin \bigcup_i \Theta_i$*,*
- *the set H consists of all sequences $(a_1, \ldots, a_\ell) \in \Sigma^*$ such that $\ell \leq n$, $a_i \in \Theta_i \cup \{\texttt{ABORT}, \perp\}$ for all i, and $a_i \in \Theta_i$ for at most $t-1$ values of $i < \ell$. $H = \Pi_{i \in N}(\Theta_i \cup \{\texttt{ABORT}, \perp\})$,*
- *the set Z of terminal histories consists of all $h = (a_1, \ldots, a_\ell) \in H$ that either (a) have length $\ell = n$ or (b) have $a_i \in \Theta_i$ for exactly t values of i, including $i = \ell$,*
- *for every history $h \in H$ of length $\ell - 1$, the next player function is defined by $P(h) = \ell$,*
- *for every non-terminal history $h = (a_1, \ldots, a_{\ell-1}) \in H \setminus Z$ and every $\theta_\ell \in \Theta_\ell$, the set of available public actions to player $\ell = P(h)$, is $A(\theta_\ell, h) = \{\theta_\ell, \texttt{ABORT}, \perp\}$,*
- *for every non-terminal history $h = (a_1, \ldots, a_{\ell-1}) \in H \setminus Z$, and any $\theta_\ell \in \Theta_\ell$, the next-message function of player $\ell = P(h)$ is defined as:*

$$m^*_\ell(\theta_\ell, a_1, \ldots, a_{\ell-1}) = \begin{cases} \theta_\ell & \text{if } a_j \in \Theta_j \quad \text{for all} \quad j < \min\{\ell, t\} \\ \texttt{ABORT} & \text{if } a_j \in \{\texttt{ABORT}, \perp\} \quad \text{for some } j < \min\{\ell, t\} \end{cases} \tag{2}$$

- *for every terminal history $h = (a_1, \ldots, a_\ell) \in Z$, and every $\theta_i \in \Theta_i$, the output function of player $i \in N$ is defined as*

$$f^*_i(\theta_i, h) = \begin{cases} g_{R_{-i} \cup \{i\}}((a_j)_{j \in S}, \theta_i) & \text{if } R_{-i} = \{j \neq i : a_j \in \Theta_j\} \text{ is of size at least } t-1 \\ \delta_0 & \text{otherwise.} \end{cases}$$

where the g_T's are the reconstruction functions from the secret-sharing scheme, and δ_0 is an arbitrary element of Δ.

A crucial property of the honest strategy is that *even players that adhere to it do not necessarily reveal their share.* This will be the case as soon as one of the first $t-1$ players aborts or deviates (by broadcasting the $\perp$ message). On the other hand, if none of the first $t-1$ players aborts or deviates, the honest strategy instructs to reveal, even if the history subsequent to the first $t-1$ rounds does contain an abort or a deviation.

Intuitively, allowing honest players to sometimes abort ensures that players who deviate from the protocol's prescribed strategy in early rounds (the first $t-1$) are penalized and unable to learn the secret themselves. On the other hand, the honest players will continue to report their share even if a deviation occurs in the t^{th} round; this ensures that the players in the first $t-1$ rounds will learn the secret in addition to the other players, and thus provides fairness.

5.4 Rational strategy for corresponding fail-stop games

We consider fail-stop games that correspond to the reconstruction protocol Π with honest players. By our hypothesis, an induced fail-stop game is a game with public actions and private outputs $\Gamma = (N, H, P, A, F, \Delta, \Theta \times \Omega, \mu \times \zeta_k, u)$.

Our goal is to describe a rational strategy for the fail-stop game. At a high-level, the strategy instructs honest players to follow the strategy prescribed by Π, whereas it instructs the rational players to deviate as as soon as they reach a point in which they can reconstruct the secret, but have the power to prevent others from learning the secret (by refusing to reveal their share). By the threshold property of the secret-sharing scheme, this occurs whenever $t-1$, but not more, values have been already broadcast. Whereas the prescribed protocol instructs the player to reveal her value (provided that no `ABORT` or $\perp$ messages have been broadcasted in any of the first $t-1$ rounds), the rational strategy will instruct the player not to reveal. This is a deviation from the prescribed strategy; the action taken by the rational player at this point will be $\perp$.

Definition 5.4 (The rational strategy)**.** *The* rational strategy *for the game Γ is the strategy vector $s = (m, f)$ that is defined as follows:*

- *For every non-terminal history $h = (a_1, \ldots, a_{\ell-1}) \in H \setminus Z$ and every $\theta_\ell \in \Theta_\ell$, the public strategy of player $\ell = P(h)$ is defined as:*

$$m_\ell(\theta_\ell, \texttt{honest}, a_1, \ldots, a_{\ell-1}) = m^*_\ell(\theta_\ell, h)$$

$$m_\ell(\theta_\ell, \texttt{rational}, a_1, \ldots, a_{\ell-1}) = \begin{cases} \perp & \text{if } (\ell \geq t) \ \& \ (a_j \in \Theta_j \text{ for all } j < t) \\ m^*_\ell(\theta_\ell, h) & \text{otherwise} \end{cases}$$

- *For every terminal history $h \in Z$, we take the private strategies $(f_1(\cdot,\cdot,h), \ldots, f_n(\cdot,\cdot,h))$ to be* any *fixed Nash equilibrium $f_{\langle h \rangle}$ of the subgame $\Gamma(s, \langle h \rangle)$, where $\langle h \rangle$ is the history h with all* `ABORT`*'s replaced by $\perp$. Such a strategy profile always exists by Nash's Theorem; as discussed earlier, any randomization needed for mixed strategies can be incorporated into the θ_i's.*[4]

The definition of the private strategies using Nash's Theorem is the only place where the rational strategy depends on the utilities in the specific fail-stop game Γ induced by Π. In case at least $t-1$ parties other than i have revealed (i.e. $R_{-i} = \{j \neq i : a_j \in \Theta_j\}$ is of size at least $t-1$), then party i's private strategy f_i can wlog be taken to be equal to the honest strategy f^*_i, which uses the reconstruction function of the secret sharing scheme (i.e. $g_{R_{-i} \cup i}$). This is because player i prefers to compute the secret correctly above all else. However, among the players for whom $|R_{-i}| < t-1$, the game $\Gamma(s, \langle h \rangle)$ may have nontrivial Nash equilibria; even though each player can only guess the secret correctly with probability $1/|\Delta|$, various strategy profiles may induce correlations among the successes of the individual players and thereby affect their utilities. In the natural case that the utility functions are *linear*, the correlations don't matter and each player can simply follow the honest strategy f^*_i, which maximizes her probability of computing the secret correctly. (By *linear* utility functions, we mean that the utility of player i is of the form $\sum_j a_{ij} c_j$ where c_j is a bit indicating whether player j computed the secret correctly, $a_{ij} < 0$ for $j \neq i$, and $a_{ii} > -\sum_{j \neq i} a_{ij}$.)

It can be seen that the strategy vector s is fail-stop admissible. This is because for every $i \in N$, and $h \in H \setminus Z$ if $m_i(\theta_i, \texttt{rational}, h) = \perp$, for some $\theta_i \in \Theta_i$ then $m_i(\theta'_i, \texttt{rational}, h) = \perp$ for all $\theta'_i \in \Theta_i$. (Note that the equilibrium notions in Theorem 5.2, which we will establish below, require that s is an equilibrium even when players may deviate to non-fail-stop strategies.)

Note that if all players follow s and there exists an honest player $i \in N$ with $i \geq t$ then t out of the n players eventually reveal their share. Thus, if we can argue that: (1) it is likely to have an honest player $i \in N$ with $i \geq t$, and (2) rational players are incentivized to follow s, we will have obtained a fair solution to the secret sharing problem (since all players are likely to learn the

[4] This definition is not circular, because $\Gamma(s, \langle h \rangle)$ depends on only the public strategy m.

secret following the protocol's execution). We start by proving that if all players follow the rational strategy s then everybody is likely to learn the secret with certainty.

Lemma 5.5 (Learning with certainty)**.** *Suppose that players' types are chosen according to $\mu \times \zeta_k$, and that all players follow the strategy vector s. Then the probability that all players know the secret with certainty in Γ is at least $1 - p(n,k)$.*

Proof. Since the players follow the rational strategy s, the first $t-1$ players will reveal their shares. So if there is a single honest player among the final $n-t+1$ players, all players will know the secret with certainty. The probability that this is not the case is

$$\left(1 - \frac{k}{n}\right)^{n-t+1} \leq e^{-k\cdot(n-t)/n},$$

as claimed. ■

6 Bayesian Subgame Perfect Equilibrium

We prove that for the rational strategy vector s is a Bayesian subgame perfect equilibrium in the extensive game with public actions and private output Γ.

Lemma 6.1 (Bayesian subgame perfect equilibrium)**.** *The rational strategy vector $s = (m, f)$ is a Bayesian subgame perfect equilibrium in the game Γ. Moreover, s is a strict Nash equilibrium in Γ.*

Proof: Our goal is to prove that the strategy vector $s = (m, f)$ is a Bayesian subgame perfect equilibrium for the game $\Gamma = (N, H, P, A, F, \Delta, \Theta \times \Omega, \mu \times \zeta, u)$. That is, we need to show that for every history $h \in H$, $s|_h$ is Nash equilibrium for the subgame $\Gamma(s, h)$.

For terminal histories h, we argue that the subgame $\Gamma(s,h)$ is equivalent to the subgame $\Gamma(s, \langle h \rangle)$, so a Nash equilibrium of the latter (which we have taken the rational strategy to be) is also a Nash equilibrium of the former. The reason is that according to equilibrium play, neither $\perp$'s nor `ABORT`'s provide any information about a player's share θ_i, and thus the distributions $(\mu \times \zeta)|_h$ and $(\mu \times \zeta)|_{\langle h \rangle}$ can be decomposed as product distributions $(\mu \times \zeta)|_h = \mu|_h \times \zeta|_h$ $(\mu \times \zeta)|_{\langle h \rangle} = \mu|_{\langle h \rangle} \times \zeta|_{\langle h \rangle}$ whose first components are equal, i.e. $\mu|_h = \mu|_{\langle h \rangle}$. The second components may differ (e.g. a $\perp$ implies that a player is rational), but are independent of the secret so do not affect the equilibria of the game.

So we focus on nonterminal histories h. Suppose that $s|_h$ is not Nash equilibrium for $\Gamma(s,h)$. That is, there is a player i and a strategy $s'_i \neq s_i|_h$ such that $u_i(\mu|_h, (s_{-i}|_h, s'_i)) > u_i(\mu|_h, s|_h)$. Without loss of generality, we may assume that player i is the first player to move in the subgame $\Gamma(s,h)$. (Otherwise, we can consider a subgame $\Gamma(s,(h,h'))$, where h' is is a history of $\Gamma(s,h)$ such that $P(h') = i$ and player i's utility increases following $s'_i(h')$ instead of $s_i|_h(h')$.)

Thus we need to argue that at nonterminal history $h = (a_1, \ldots, a_{\ell-1})$, player ℓ cannot increase her utility (in $\Gamma(s,h)$) by deviating from the rational strategy in the public action she takes. Recall that if $\omega_\ell = \texttt{honest}$, player ℓ has only one action to choose from (as determined by $m^*_\ell(\theta_\ell, h)$). Thus it suffices to focus on the case that $\omega_\ell = \texttt{rational}$, in which player ℓ has two actions to choose from (either the action determined by the honest strategy $m^*_\ell(\theta_\ell, h)$ or $\perp$).

Let $\mathsf{rev}(h)$ denote the number of *reveal* actions in h. That is, $\mathsf{rev}(h)$ is the number of $j \leq \ell - 1$ for which $a_j = \theta_j$. Since the public portion of the protocol and game ends as soon as there are t reveals, the nonterminal history h has $\mathsf{rev}(h) \leq t-1$. Let $\bar{\mu} = \mu \times \zeta$ and recall the way in which $\bar{\mu}|_h$ is defined (see Definition 2.3).

Our analysis distinguishes between three cases, depending on whether the rational strategy instructs player ℓ to set θ_ℓ, ABORT, or $\perp$. We begin with the cases that the strategy instructs player ℓ to send $\perp$, i.e. when $\ell \geq t$ and all players from 1 to $t-1$ have revealed their share. In such a case, the alternative strategy available to player ℓ is to reveal her share (because that is the honest strategy).

Claim 6.2. *If $\ell \geq t$ and $a_j \in \Theta_j$ for all $j < t$, then $u_\ell|_h(\overline{\mu}|_h, (s_{-\ell}, \perp)) > u_\ell|_h(\overline{\mu}|_h, (s_{-\ell}|_h, \theta_\ell))$.*

Proof of Claim: At the start of subgame $\Gamma(s,h)$, player ℓ already knows the secret with certainty, but at least one other player does not (because $\mathsf{rev}(h) < t$). If she sends θ_ℓ, then all players will know the secret with certainty. On the other hand, if she sends $\perp$, there is a nonzero probability that some player will fail to compute the secret. The reason is that there is a nonzero probability that all future players will be rational and thus will also send $\perp$, revealing no additional information about the secret. ■

Next, we consider the case that the rational strategy instructs player ℓ to send ABORT. This occurs exactly when the honest strategy is to send ABORT, namely when one of the first $t-1$ players has not revealed her share.

Claim 6.3. *If there exists a $j < t$ such that $a_j \in \{\texttt{ABORT}, \perp\}$, then $u_\ell|_h(\overline{\mu}|_h, (s_{-\ell}, \texttt{ABORT})) = u_\ell|_h(\overline{\mu}|_h, (s_{-\ell}|_h, \perp))$*

Proof of Claim: Whether player ℓ sends $\perp$ or ABORT, the rational strategy will instruct all future players to send ABORT. Thus the terminal histories reached in both cases will have identical "reduced" forms $\langle h \rangle$, and player ℓ will receive whatever her expected utility is in the game $\Gamma(s, \langle h \rangle)$. ■

Finally, we consider the case that the rational strategy instructs player ℓ to reveal her share. This occurs when $\ell < t$ and all previous players have revealed their shares. Our analysis of this case utilizes the existence of honest players, and is the place where we need Condition 1.

Claim 6.4. *If $\ell < t$ and $a_j \in\in \Theta_j$ for all $j < \ell$, then $u_\ell|_h(\overline{\mu}|_h, (s_{-\ell}, \theta_\ell)) > u_\ell|_h(\overline{\mu}|_h, (s_{-\ell}|_h, \perp))$.*

Proof of Claim: We estimate the expected utility that player ℓ receives for revealing her share θ_ℓ versus sending $\perp$.

If player ℓ reveals her share, then the rational strategy will instruct the remaining players $\ell+1, \ldots, t-1$ to also reveal their shares. Thus if there is at least one honest party among parties $t, \ldots, n$, then t shares will be revealed and everyone will compute the secret correctly, so player ℓ receives utility $u_\ell(o_N)$. This occurs with probability at least $1 - p(n,k)$. Otherwise, player ℓ receives utility at least $u_\ell(o_{N\setminus\{i\}})$, since $o_{N\setminus\{i\}}$ is the worst outcome for player i. So by revealing her share, player ℓ receives utility at least $(1-p(n,k)) \cdot u_\ell(o_N) + p(n,k) \cdot u_\ell(o_{N\setminus\{i\}})$.

If player ℓ does not reveal her share, then the rational strategy will instruct the remaining players to send ABORT. By the secrecy property of the secret-sharing scheme, player ℓ will then be able to compute the secret correctly with probability only $1/|\Delta|$. Thus she will receive expected utility at most $(1/|\Delta|) \cdot u_\ell(\{i\}) + (1 - 1/|\Delta|) \cdot u_\ell(o_\emptyset)$.

Therefore, player ℓ has a strict preference to reveal her share provided that:

$$(1-p(n,k)) \cdot u_\ell(o_N) + p(n,k) \cdot u_\ell(o_{N\setminus\{i\}}) > (1/|\Delta|) \cdot u_\ell(o_{\{}i\}) + (1 - 1/|\Delta|) \cdot u_\ell(o_\emptyset).$$

This is equivalent to Condition 1. ■

■

We have shown that the rational strategy s is a Bayesian subgame perfect equilibrium. To see that it is also a strict Nash equilibrium, we observe that the only public cases where a player may be indifferent between sending the message specified by s and an alternative strategy is when s specifies sending `ABORT` (Claim 6.3), and this occurs with zero probability in equilibrium.

Remark 6.1. *As discussed earlier, the honest-player distribution ζ_k in Theorem 5.2 models the situation where players have no a priori information about which players are honest or rational, except the probability (k/n) with which an individual player is honest. The independence between different players in ζ_k is not essential to the proof; it applies to any distribution in which there is a high probability of an honest player among $t, \ldots, n$, such as the uniform distribution on vectors $\omega \in \{\texttt{honest}, \texttt{rational}\}^n$ in which exactly k components are* `honest`*. An opposite extreme is the case where the set of honest players is not random, but is instead an arbitrary fixed set of k players, known to all. Our protocol can be modified to handle this case by first randomly permutating the order in which players speak; this guarantees that there will be an honest player among $t, \ldots, n$ with high probability. Now the rational strategy would have the first $t-1$ players determine whether the permutation is 'good', reveal their share if it is, and send $\perp$ otherwise. This analysis can be found in the preliminary version of our paper [**?**], where it is also shown to be coalition-proof in the sense of Bernheim et al. [4]." By combining these ideas, it may be possible to handle arbitrary distributions on the set of honest players (provided there are at least roughly k honest players with high probability).*

7 Acknowledgements

We thank Drew Fudenberg, Jonathan Katz, Silvio Micali, Peter Bro Miltersen, and Moni Naor for helpful discussions.

References

[1] *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing, 2-4 May 1988, Chicago, Illinois, USA*. ACM, 1988.

[2] I. Abraham, D. Dolev, R. Gonen, and J. Y. Halpern. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. In E. Ruppert and D. Malkhi, editors, *PODC*, pages 53–62. ACM, 2006.

[3] A. S. Aiyer, L. Alvisi, A. Clement, M. Dahlin, J.-P. Martin, and C. Porth. Bar fault tolerance for cooperative services. In A. Herbert and K. P. Birman, editors, *SOSP*, pages 45–58. ACM, 2005.

[4] B. P. B. Douglas Bernheim and M. D. Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.

[5] M. Babaioff, R. Lavi, and E. Pavlov. Mechanism design for single-value domains. In *Proc. Nat. Conf. on Artificial Intelligence, AAAI05*, 2005.

[6] M. Ben-Or, S. Goldwasser, and A. Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In *STOC* [1], pages 1–10.

[7] G. Blakely. Safeguarding cryptographic keys. In *AFIPS*, volume 48, page 313, 1979.

[8] D. Boneh and M. Naor. Timed commitments. In *Proc. CRYPTO 2000*, pages 236–254, 2000.

[9] D. Chaum, C. Crépeau, and I. Damgård. Multiparty unconditionally secure protocols (extended abstract). In *STOC* [1], pages 11–19.

[10] B. Chor, S. Goldwasser, S. Micali, and B. Awerbuch. Verifiable secret sharing and achieving simultaneity in the presence of faults (extended abstract). In *FOCS*, pages 383–395. IEEE, 1985.

[11] R. Cleve. Limits on the security of coin flips when half the processors are faulty (extended abstract). In *STOC*, pages 364–369. ACM, 1986.

[12] S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. *Commun. ACM*, 28(6):637–647, 1985.

[13] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based mechanism for lowest-cost routing. In *Proceedings of the 2002 ACM Symposium on Principles of Distributed Computing*, pages 173–182, 2002.

[14] J. Feigenbaum, C. H. Papadimitriou, and S. Shenker. Sharing the cost of multicast transmissions. *Journal of Computer and System Sciences*, 63:21–41, 2001.

[15] J. Feigenbaum and S. Shenker. Distributed Algorithmic Mechanism Design: Recent Results and Future Directions. In *Proceedings of the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pages 1–13, 2002.

[16] J. A. Garay and M. Jakobsson. Timed release of standard digital signatures. In *Proc. Financial Cryptography 2002*, pages 168–182, 2002.

[17] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game or a completeness theorem for protocols with honest majority. In *STOC*, pages 218–229. ACM, 1987.

[18] S. D. Gordon and J. Katz. Rational secret sharing, revisited. In R. D. Prisco and M. Yung, editors, *SCN*, volume 4116 of *Lecture Notes in Computer Science*, pages 229–241. Springer, 2006.

[19] J. Y. Halpern and V. Teague. Rational secret sharing and multiparty computation: extended abstract. In L. Babai, editor, *STOC*, pages 623–632. ACM, 2004.

[20] S. Izmalkov, S. Micali, and M. Lepinski. Rational secure computation and ideal mechanism design. In *FOCS*, pages 585–595. IEEE Computer Society, 2005.

[21] G. Kol and M. Naor. Games for exchanging information. manuscript. 2007.

[22] D. Lehmann, L. I. O'Callaghan, and Y. Shoham. Truth revelation in approximately efficient combinatorial auctions. *Journal of the ACM*, 49(5):577–602, September 2002.

[23] M. Lepinski, S. Micali, C. Peikert, and A. Shelat. Completely fair sfe and coalition-safe cheap talk. In S. Chaudhuri and S. Kutten, editors, *PODC*, pages 1–10. ACM, 2004.

[24] M. Lepinski, S. Micali, and A. Shelat. Collusion-free protocols. In H. N. Gabow and R. Fagin, editors, *STOC*, pages 543–552. ACM, 2005.

[25] A. Lysyanskaya and N. Triandopoulos. Rationality and adversarial behavior in multi-party computation. In C. Dwork, editor, *CRYPTO*, volume 4117 of *Lecture Notes in Computer Science*, pages 180–197. Springer, 2006.

[26] R. McGrew, R. Porter, and Y. Shoham. Towards a general theory of non-cooperative computation. In J. Y. Halpern and M. Tennenholtz, editors, *TARK*, pages 59–71. ACM, 2003.

[27] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35:166–196, 2001.

[28] D. C. Parkes and J. Shneidman. Distributed implementations of Vickrey-Clarke-Groves mechanisms. In *Proc. 3rd Int. Joint Conf. on Autonomous Agents and Multi Agent Systems*, pages 261–268, 2004.

[29] A. Petcu, B. Faltings, and D. Parkes. M-dpop: Faithful distributed implementation of efficient social choice problems. In *AAMAS'06 - Autonomous Agents and Multiagent Systems*, pages 1397–1404, Hakodate, Japan, May 2006.

[30] B. Pinkas. Fair secure two-party computation. In *Proc. EUROCRYPT 2003*, pages 87–105, 2003.

[31] T. Rabin and M. Ben-Or. Verifiable secret sharing and multiparty protocols with honest majority (extended abstract). In *STOC*, pages 73–85. ACM, 1989.

[32] A. Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979.

[33] J. Shneidman and D. C. Parkes. Specification faithfulness in networks with rational nodes. In *Proc. 23rd ACM Symp. on Principles of Distributed Computing (PODC'04)*, St. John's, Canada, 2004.

[34] Y. Shoham and M. Tennenholtz. Non-cooperative computation: Boolean functions with correctness and exclusivity. *Theor. Comput. Sci.*, 343(1-2):97–113, 2005.

[35] A. C.-C. Yao. How to generate and exchange secrets (extended abstract). In *FOCS*, pages 162–167. IEEE, 1986.