Renegotiation-Safe Protocols

Rafael Pass*
rafael@cornell.edu

abhi shelat†
abhi@virginia.edu

August 19, 2010

Abstract

We consider a model of renegotiation in extensive-form games: when it is player i's turn to move, i can "renegotiate" the equilibrium by suggesting new strategies for all players for the remainder of the game. This renegotiation is successful if it improves i's utility, and cannot itself be renegotiated at a later round in the game. Although not all finite games have *renegotiation-safe* strategies, natural classes of games do.

We argue that renegotiation-safety captures rationality in the context of cryptographic protocols in a more meaningful way than traditional solution concepts. We also present protocols for the task of secret sharing that are renegotiation-safe assuming the existence of two non-negotiating players; additionally, we show that such protocols require the existence of at least one non-negotiating player.

*Pass is supported in part by a Microsoft Research Faculty Fellowship, NSF CAREER Award CCF-0746990, AFOSR Award FA9550-08-1-0197, and BSF Grant 2006317.

†shelat is supported in part by a Microsoft Research Faculty Fellowship and an NSF CAREER Award CNS-0845811.

1 Introduction

A long-term goal in both game theory and cryptography is to understand the role of communication and bounded rationality on the outcome of interactive processes. Recent efforts have been made to introduce elements of the game theoretic model to the cryptographic model of interaction. One prevailing tool has been the notion of Nash equilibrium and its various extensions. The weaknesses of a Nash equilibrium, however, are well understood. Nash equilibria may rely on "empty threats", and they only consider single-player deviations. Thus it may be ill-suited to a game that inherently enables player collusion.

More broadly, the equilibrium notion seems problematic in the context of protocol design. The goal from the onset is to design a protocol with predictable outcomes that satisfy well-defined optimality properties. But even the strongest equilibrium concepts do not appear to provide such guarantees. For instance, the notion of a sequential equilibrium [KW82] eliminates empty threat by requiring that the equilibrium strategy specifies beliefs under which the threats become credible; however, as is well-understood, such beliefs are sometimes themselves not "reasonable" (see e.g., the classic Kohlberg-Mertens [KM86] example of a "poor" sequential equilibrium in Fig. 7). More generally, (and again, as is well-understood; see e.g., [OR94]), in classic equilibrium concepts, a deviating player is expected to continue playing honestly, even after deviating. But the whole point of deviating might have exactly been to signal a future deviation.

Thus, protocol designers often have in mind a much stronger goal than equilibrium. The classic approach in the context of mechanism design is to require that the desired outcome that can be implemented in *dominant* strategies. But this approach is often overly restrictive; see e.g., the Myerson-Satterthwaite theorem [MS83]. The problem becomes more troublesome in the context of communication protocols—dominant strategy implementations do not support "collaborative protocols": Consider a game with two players; both players prefer to "talk" if the other player does, but prefer not to talk if the other player does not. Clearly, in such a situation we would expect the players to talk, but talking is not a dominant strategy.

In this paper, we consider a less restrictive, but meaningful notion of rationality for communication games. The core new idea is to adapt the idea of *renegotiations* introduced by Farrell in 1983 [Far83] in the context of mechanism design. Briefly, we allow players the extra ability to "renegotiate" the entire strategy profile for all players when it is their turn to make a move as long as (1) the renegotiation benefits them, and (2) the new strategy profile cannot later be renegotiated. This new notion avoids weaknesses of Nash equilibrium (and for example, prevents a notion of empty threats) and also applies to both games of incomplete information and games with computationally bounded agents *without* resorting to the assumption that deviating players will continue playing honestly.

Renegotiation is of course not a new concept in the game theory literature. See, for instance, Farell [Far83], Farell and Maskin [FM87], Bernheim and Ray [DBR89], van Damme [vD89], Blume [Blu87], Bernheim and Whinston [BW87], Benoit and Krishna [BK93], Cave [Cav87], Asheim [Ash91], DeMarzo [DeM88], Bergin and Macleod [BM89]. However, as far as we know, these earlier notions only apply to games of complete information; none consider the setting of computationally bounded agents. Thus, none of them seem appropriate for modeling cryptographic interactions, where player have secret inputs (and thus the game has incomplete information) and often are computationally bounded. Our notion attempts to rectify this. To explain it, let us first provide a brief survey of prior work in the game theory literature. The notion informally put forth by Farrell considers a T-fold repetition of a single shot game G. In other words, at every stage, players are faced with the *same* decision, players make a simultaneous move, and receive a payoff for that stage. For the case of two-players, this notion is equivalent to coalition-proof Nash equilibrium introduced by Bernhiem, Peleg and Whinston, and also Pareto-perfect equilibrium introduced by Bernheim and Ray. As expressed by Bernheim and Ray, the intuition behind the notion is that:

> *. . . we require that an equilibrium [does] not prescribe any course of action in any subgame that players would jointly wish to renegotiate, given the restriction that any alternative must*

itself be invulnerable to subsequent deviations and renegotiation.

Ferreira [Fer95] further develops the connection between renegotiation and coalitions by through his notion of Communication-proof equilibria. First, he extends the renegotiation notion to general extensive form games (instead of T-fold repetitions) with perfect information. In doing so, he requires the renegotiation-safety property to hold *at every history* of the game. Second, he allows a coalition (of the players that move at a given round) to renegotiate (instead of requiring that all the players *jointly* wish to renegotiate).

Our notion differs from that of Ferriera in these last two points. In order to apply our notion to games of incomplete information (and also for the case of computationally-bounded agents), we do not require the property to hold at every history; instead we require there is no "round" r in the game such that a player acting in this round has incentives to renegotiate the equilibrium strategy.[1] Second, to simplify our discussion, we only consider renegotiations that are proposed by a single player for the next round. This decision is driven by our eventual goal of analyzing games in which only one player broadcasts a message at a time. We note that Asheim [Ash97] proposes a similar notion when considering time consistency of a single player who plans for the future by making decisions at every time step. However, just as earlier notions, his notion is for perfect-information games, and considers renegotiation at every history.

1.1 Renegotiation-Safe Strategies

In standard equilibrium, a player is assumed to follow the equilibrium strategy. A player who deviates once, is also assumed to continue following the equilibrium strategy *after* the deviation. As mentioned, this interpretation of the rational model has been well-discussed and criticized. For example, it seems reasonable to also believe that a player who has deviated may continue to deviate (after all, the player might have deviatied for a reason). Our, as well as earlier, notions of renegotiation (and explicitly [Fer95]) handle this shortcoming of the standard equilibrium model by allowing a deviating player to not only *signal* his deviation, but to suggest an entirely new profile of strategies for the players.

In our model, renegotiation is only allowed by the players who make moves at round r. Alternatively, we may allow any player to suggest a renegotiation at any round. We believe that the structure of the game is important, and that only players who can *signal* the formation of a coalition have the strategic power to renegotiate. Thus, we preserve this power in our model. In any case, it is easy to add dummy messages to a mechanism or protocol to give the strategic ability to renegotiate to every player if that is desired.

As mentioned, earlier notions of renegotiation aim to capture the intuition that there does not exist a *history* in the game where a player wants to renegotiate. Our notion instead aims to capture the intuition that, *a-priori*, before having started to play the game, no player can expect to improve utility from renegotiation. Roughly speaking, *renegotiation safety (RS)* captures the intuition that, *a-priori*, no the player can improve her utility by, potentially at a future round, suggesting a renegotiation of strategies, such that this new renegotiated strategy is self is stable to further renegotiations. Technically, this means that we are only evaluating the utility of a renegotiation from the "root node" (and thus, renegotiations at histories that *a priori* have probability 0 of being reached are considered useless). As mentioned, this is the property that allows us to deal with incomplete information, and as we shall see shortly, also computationally bounded agents.

[1] The problem with quantifying over every history in a game of incomplete information is that doing so requires specifying player beliefs at each such history. It is conceivable that Ferriera's notion could be extended to games of incomplete information in analogy with the notion of sequential equilibrium. But then we would be inheriting the problems of sequential equilibrium, which is exactly what we aim to circumvent. Additionally, as we discuss later on, history-based equilibrium concepts become less meaningful in the context of computationally-bounded agents.

1.2 Renegotiation by Computationally-bounded Agents

Just like the notion of Nash equilibria, our notion of RS is also meaningful when restricting the players to be computationally bounded—i.e., strategies that can be implemented by computationally-restricted Turing Machines. In contrast, solution concepts such as subgame perfection (or sequential equilibrium), or earlier notions of renegotiation proofness, become less meaningful in such situations. The reason is that these solution concepts require "optimality" on every history of the game. It is well-known that for natural games in the context of cryptographic protocols, any *a priori* optimal strategy can never be optimal at every history (for instance, given a history where a public key has been announced, there is always a computationally-bounded Turing Machine that "knows" the secret key associated with the public key). Since our notion of renegotiation only consider the *a priori* gain of a player wanting to renegotiate at a (potentially) later round, we circumvent this problem. In particular, our definition remains meaningful no matter whether we allow agents to choose any strategy (as in the traditional game-theoretic model) or whether we limit them to a set of computationally-bounded strategies.

We mention that recent work consider the question of redefining sequential equilibrium with respect to computationally bounded agents: Halpern and Pass [HP10a] show that in a model where computation is costly, a generalization of the traditional notion may again become meaningful; the very recent work of Gradwohl, Livne, and Rosen [GLR10] on the other hand, present a weakening of sequential equilibrium that can be meaningfully achieved; the [GLR10] notion also seems to considers an a priori notion of deviation. These works however do not consider renegotiation.

1.3 Public Renegotiation versus Secret Deviation

It is instructive to compare RS and NE. In a NE, a player does not want to deviate—i.e., unilaterally changes his strategy, but saying nothing else to the other players—if everyone else sticks to their strategy. In an RS, a renegotiation explicitly—and publicly—specifies what strategies the other players use. Thus, the type of deviations considered by the two notions are quite different: NE considers deviations where the deviator does not necessarily announce its deviation (or the reasons for it); RS instead considers deviations where the deviator explicitly announces its deviation and the reasons for it (and how everyone else should proceed in a way that is safe for them). Indeed, not all RS strategies are NE (see the rightmost game in Figure 4 for an example). This might look weak: if a strategy is not a NE, there exists some player i that can increase her utility by unilaterally deviating; if so, why shouldn't she? If the strategy is RS, player i will be concerned that if she deviates, *and this deviation is detected by the other players*, then some other player j might later renegotiate the strategy in a way that actually decreases i's utility. Thus, in a model where all deviations can be detected, RS alone seems to provide strong guarantees. In models where detections might not be as easily detected (e.g., in bayesian games, or games with computationally bounded agents), it seems safer to ask for stability against both public renegotiaions and secret deviations; that is to consider both RS and some notion of stability against potentially secret deviations (e.g., NE or sequential equilibrium).

1.4 Our results

We first provide a simple example of a 2-round 2-player games without any RS strategies. However, although, RS strategies do not always exists, we show that they do exist for a natural class of games. Roughly speaking, our existence theorem show that all finite *sequential games*—where at every round in the game, except the last one, at most one player moves (in the last round, many players might simultaneously move), and all the moves are perfectly observed by all the players—have RS strategies; furthermore, every such game has an SPE that is RS.

Additionally, as we argue, RS provides insight into the design of cryptographic protocols. As a test-bed, we apply our solution for the design of "rational" cryptographic protocols to the tasks of *secret sharing*

when players have strictly competitive utilities—namely, players want to get the correct output, and if they do, they prefer that as few other players as possible get it (a setting first studied in [HT04], and more recently in e.g., [GK06, ADGH06, KN08a, KN08b, IML05, MS09, OPRV09]; see Appendix A for a brief overview of this literature). We first argue that whereas traditional solution concepts do not seem to appropriately capture "rationality" in the context of secret sharing protocols (in that they do not rule out protocols that we, intuitively, would deem "bad"), combining RS and NE leads to a better behaved notion. We then show that there are no fixed-round RS-secret sharing reconstruction protocols. Finally, we apply this infeasibility result to establish a secret sharing reconstruction protocol that is both a NE and satisfies a notion of RS, assuming the existence of two players that are not willing to renegotiate. To illustrate how our model applies to the setting of computationally-bounded agents, we also show that if furthermore assuming that players are computationally-bounded, then our protocol remains RS even if our protocol is implemented using cryptographic primitives that only are "computationally secure".

2 Defining Renegotiation Safe Strategies

2.1 Preliminaries

Bayesian games with publicly observed actions We restrict out attention to games with a fixed-schedule, i.e., for any two histories h_1, h_2 of equal length, the players who move at those histories are the same. Additionally, we restrict attention only to games where all moves are publicly observable. More formally, a bayesian game of publicly observed actions Γ is a 9-tuple consisting of the following:

1. A set of players $[n] = \{1, \ldots, n\}$,
2. A set of histories H and a subset $Z \subseteq H$ of terminal histories; all terminal histories have the same length (we refer to this as the length of the game, or the number of rounds in the game),
3. a player function $P(h)$ which maps a history to a set of players who make the next move, such that for any two histories h_1, h_2 of equal length $P(h_1) = P(h_2)$.
4. a set of actions $A = A_1 \times \cdots \times A_n$ where A_i is the set of actions for player i,
5. a set of types Θ_i for $i \in \{0, \ldots, n\}$; $\theta_0 \in \Theta_0$ should be thought of as the type of "nature" and $\theta_i \in \Theta_i$ where $i > 0$ is the type of player i,
6. a distribution μ over Θ where $\Theta = \Theta_0 \times \cdots \times \Theta_n$,
7. a utility function $u_i(\vec{\theta}, h)$ for each player that maps to the interval $[0, 1]$,
8. and a compact set of strategies for the players, S.

The model of a bayesian games captures the following process: μ is sampled to produce a type for nature θ_0 and player types $\theta_1, \ldots, \theta_n$. The type θ_i is given to player i. This is followed by a sequence of actions that are visible to all players: After any history $h \in H$, the set of players $i \in P(h)$ chooses actions from their respective action sets $A_i(h)$. This choice determines the next actions of the players and so on until a terminal history $h \in Z$ is reached. The utility of player i is then determined to be the value $u_i(\vec{\theta}, h)$. Per standard convention, θ_{-i} denotes the profile of types for all players but i and a_{-i} denotes the actions of all players but i.

Strategies The set S traditionally consists of the full set of mixed behavioral strategy; this is clearly a compact set. Unless, mentioned otherwise, we always consider this case. We will also consider the case when S consists of T-bounded strategies (i.e., strategies that can be implemented by a circuit of size at most T); this set of T-bounded strategies is finite and thus also compact.

Sequential bayesian games A *sequential bayesian game* is a bayesian game where $|P(h)| = 1$ for every history of length less than or equal to $N - 1$, where N is the number of rounds in the game; that is, only one player makes a move at a time, except possibly in the last round where any number of players move. We refer to sequential bayesian games in which the type information for each player is empty (i.e., the game is non-bayesian) as simply *sequential games*. (Note that a *perfect information* game is a sequential game where also in the last round only one player make a move.) As usual, a game is finite if the type set and the actions sets are finite.

Definition 1 (NE)**.** A strategy σ is a Nash Equilibrium (NE) if for each player i, and each strategy σ_i' for player i, it holds that $u_i(\sigma) \geq u_i(\sigma_i', \sigma_{-i})$. The NE is *strict* when the inequality is strict.

Subgames Given a bayesian extensive-form game G, a history h, and a strategy σ that reaches h with positive probability (when the types are selected according to the type distribution in G), let $G_\sigma(h)$ denote the game obtained by fixing the history in G to h and letting the type distribution be determined by sampling a type profile $\vec{t}$ according to the type distribution in G, but conditioned on $\sigma(\vec{t})$ yielding the history h. Note that if G is a complete information game (i.e., non-bayesian), then for every strategy σ that reaches h, $G_\sigma(h) = G'_\sigma(h)$; in this case, we let $G(h)$ denote this unique game obtained by simply fixing the history to h in G.

Definition 2 (SPE)**.** A strategy σ for a sequential game is a subgame perfect equilibrium (SPE) if for every non-terminal history h, the strategy profile σ is a NE for $G(h)$.

2.2 Renegotiation Safe Profiles

We now turn to defining renegotiation-safety. Roughly speaking, a strategy is *renegotiation safe* if, *a-priori*, no player can improve its utility by at any future round r suggesting a *renegotiation* of strategies that is not susceptible to future renegotiations.

We use the notation $[\sigma, r, \sigma]$ to denote the strategy that consists of following σ for the first r rounds and then following σ' for the remainder of the game. We apply the notation to both player strategies and profiles of strategies. Furthermore, we let $[\sigma, r, \sigma]_i = ([\sigma_i, r, \sigma_i'], [\sigma_{-i}, r + 1, \sigma'_{-i}])$ to denote the strategy where player i switches from σ to σ' in round r and all other players switch in round $r + 1$.

Definition 3 (Renegotiation safety)**.** Strategy profile σ is renegotiation-safe (RS) at round r in the game G with utility function u if r is a valid round in the game, and for every player i that makes a move at round r, there is no σ' such that

1. $u_i([\sigma, r, \sigma']_i) > u_i(\sigma)$, and
2. $([\sigma, r, \sigma']_i)$ is RS for every valid round $r' > r$.

A strategy profile is RS from round r if it is RS at every round $r' \geq r$. A strategy profile is RS if it is RS from round 1.

Remark 1. Note that in rounds r where only one player moves $[\sigma, r, \sigma]_i = [\sigma, r, \sigma]$. In games with simultaneous move, the fact that we only let player i renegotiate to $[\sigma, r, \sigma]_i$ (and not $[\sigma, r, \sigma]$) means that in round r player i can only change his strategy but not the strategy of others that move in that round; intuitively, the other players strategy can only be changed after they have observed player i move in round r, and thus their new strategies are only applied to round $r + 1$.

Remark 2. Our definition of RS captures a situation where a player publicly announces its renegotiation; both the fact that the renegotiation took place and the new renegotiated strategies are common knowledge

among the players. Technically speaking, strategies are thus probabilistic functions from public histories of actions (as classically) and public histories of renegotiations, to actions. However, without loss of generality, we can assume that the renegotiated strategy σ' has the history of renegotiations "hard-coded" and thus only treat it as a classic strategy (from histories of actions to actions). However, later on, in Section 5.1, when we consider non-negotiating players, this extra generality will be useful.

Remark 3. As already mentioned, the main differences between RS and communication-proof equilibria (CPE) of Ferreira [Fer95] are (1) we only consider renegotiations, whereas CPE considers both *deviations* and renegotiations, and (2) we only consider *a-priori* renegotiation, whereas CPE consider renegotiation (and deviation) at every history of the game. Furthermore, we only consider renegotiations that are proposed by a *single player* for the next round. In contrast, CPE consider also renegotiations and deviations by coalitions (that themselves are stable to deviations by subcoalitions) of the players that can move in a certain round. At first sight, RS thus seems weaker that CPE. But, the fact that we weaken the definition of a renegotiation makes the stability condition weaker, and thus more renegotiations are considered stable; this, in turn, strengthens the definition, and thus RS is seemingly incomparable to CPE.

Remark 4. Note that in our definition of RS, we require that any renegotiation by a player i is stable to all later renegotiations, even those by the *same* player i. The reason for this is the following. If i suggests a renegotiation at round r that is susceptible to a later renegotiation by i at round r', then a player j, moving at round $r < r'' < r'$ will not be convinced by the stability of the renegotiation (indeed, he will believe that i will renegotiate at round r' and thus might not be willing to participate in the original renegotiation.

In analogy with the notion of an ϵ-NE, we say that σ is an ϵ-RS if it is RS except that the right-hand side of the inequality in condition (1) is replaced by $u_i(\sigma) + \epsilon$. In other words, any renegotiation must improve the utility of the party that proposes the renegotiation by at least ϵ. One way to think about ϵ-RS is that in a model where renegotiations cost at least ϵ (see e.g., the costly-computation model of [HP10a]), no player actually prefers to renegotiate.

Remark 5. An alternative way of defining ϵ-RS would be to additionally "weaken" condition (2) to require that the renegotiated strategy is ϵ-RS (instead of RS). The problem is that this change might not weaken the definition: In particular, an RS strategy might not be an ϵ-RS under such a definition. The reason for this is that since we weakened condition 1) more strategies are considered ϵ-RS at the last round, which in turn means that more renegotiations are considered stable at the next to last round, etc.

It directly follows from the definition that for normal-form games, RS and NE coincide.

Fact 1. *For every (Bayesian) normal-form game G, a strategy profile σ is (ϵ-)RS if and only if it is an (ϵ-)NE.*

However, as we shall see, for extensive-form games RS and NE are incomparable; in fact, even RS and subgame perfection/sequential equilibrium are incomparable.

Note that our definition of RS does not require that there are no renegotiations at any partial history (as previous definitions of renegotiation proof-ness). However, just as NE strategies are optimal at each history that is on the equilibrium path, we also have that an RS strategy σ does not have any profitable renegotiations at histories that are reached with positive probability by σ.

Lemma 1. *Let σ be a strategy profile in a bayesian extensive-form game G. Then σ is RS from round $r+1$ iff $\sigma(h)$ is RS in $G_\sigma(h)$ for every r-round history h that is reached with positive probability by σ.*

Proof: We show the lemma by induction on the number of rounds remaining in the game. More precisely, we show the following claim by induction on s:

> for every n-round game G, every $r \geq n - s - 1$, σ is RS from round $r+1$ iff σ is RS in $G_\sigma(h)$ for every r-round history h that is reached with positive probability by σ.

Base case: s=0. We need to show that σ is RS from round n iff σ is RS in $G_\sigma(h)$ for every round r history h that is reached with positive probability by σ. Since $G_\sigma(h)$ is a normal-form game, by Fact 1 we have that $\sigma(h)$ being RS in $G_\sigma(h)$ is equivalent to $\sigma(h)$ being a NE. Clearly, if $\sigma(h)$ is not a NE for some h that is reached with positive probability, then we have a renegotiation also at round n in G (and this renegotiation is trivially stable, because n is the last round). On the other hand, if σ has a profitable renegotiation at round n, then there must exists some history h (that is reached with positive probability) such that $\sigma(h)$ is not a NE (or else the renegotiation would not be profitable).

Inductive step: Assume the claim is true for every $s' \leq s-1$. We show that it is true also for s.

We start by showing the forward direction. Let σ be RS from round $r+1$ in an n-round game G where $r \geq n-s-1$. We show that $\sigma(h)$ is RS in $G_\sigma(h)$ for every r-round history h that is reached with positive probability. Assume for contradiction that there exists an r-round history h that is reached with positive probability by σ (when the types are selected according to the type distribution), such that $\sigma(h)$ is not RS in $G_\sigma(h)$. That is, there exists a player i and a stable renegotiation σ' at round t that is profitable for i. But in that case, renegotiating from σ to $\hat{\sigma}$ in round $r+t$—where $\hat{\sigma}$ is identical to σ but on histories extending h, it instead plays σ'—is also a profitable renegotiation in G. It only remains to argue that $\hat{\sigma}$ is RS at every round $r' > r+t$. If not, there exists a stable renegotiation τ at some round $r' > r+t$. By the induction hypothesis (we can apply it because $r' > n-s$ since $r' > r+t \geq r+1$), there thus exist some $(r'-1)$-round history h' such that $\hat{\sigma}$ is not RS in $G_{\hat{\sigma}}(h')$. Applying the induction hypothesis again (and the fact that σ is RS from round $r' > r+t \geq r+1$) we have that h' must extend h (since unless h is reached $\hat{\sigma}$ is the same as σ). But, if h' extends h, by the induction hypothesis, this contradict that σ' is RS at round t in $G_\sigma(h)$; note that we here can rely on the induction hypothesis since the number of rounds in $G_\sigma(h)$ is $m = n-r$, and $t \geq m-s$ (since by assumption $r \geq n-s-1$ and $t \geq 1$).

We move on to show the converse. Assume that σ is not RS from round $r+1$ in some n-round game G, where $r \geq n-s-1$; that is, there exists a profitable stable renegotiation σ' at some round $r' > r+1$. That means there exist some r-round history h that is reached with positive probability by σ, such that conditioned on reaching h, σ' is still a profitable renegotiation. But then $\sigma'(h)$ is a profitable $(r'-r)$-round renegotiation from σ in $G_\sigma(h)$. As before, it only remains to show that σ' is stable in $G_\sigma(h)$. Assume there exists some stable renegotiation τ from $\sigma'(h)$ in $G_\sigma(h)$ at round t. Consider $\hat{\tau}$ that is equal to σ', but on histories extending h, it instead plays τ. Clearly, since h is reached with positive probability, $\hat{\tau}$ is a profitable renegotiation from σ' at round $r'+t$ in G. We claim that $\hat{\tau}$ is also RS from round $r'+t+1$, which yields a contradiction. If not, then by the induction hypothesis there exists some renegotiation at some history h'. First, note that h' must extend h; this follows by the induction hypothesis and the fact that σ' is RS from $r'+1$ and τ' is identical to σ' on histories that do not pass through h. On the other hand, h' cannot extend h, since if it did, then (again applying the induction hypothesis) τ could not be stable in $G_\sigma(h)$. □

As a corollary, we have:

Lemma 2. *If σ is an RS in G then for all history h that is reached by σ, $\sigma(h)$ is RS in $G_\sigma(h)$.*

Proof: If σ is RS in G, then it is RS from every round in G and thus the lemma directly follows by Lemma 1. □.

Remark 6. It is worthwhile to note that Lemma 1 and 2 do not hold if we consider restricted strategy spaces (e.g., T-bounded strategies). Recall that in the proof of Lemma 1, we are required to "paste" together two strategies and this new strategy might no longer be in the restricted strategy space. Indeed, it is easy to come up with counter examples for the case of T-bounded strategies; for instance, if the public history contains an encrypted message (using a public-key cryptosystem) which, if broken, yields high utility. Conditional on every fixed history, there is a simple T-bounded strategy that outputs the decryption (which is unique);

but there might not exist a T-bounded strategy that breaks a random encryption (indeed, if the encryption scheme is "secure" against T-bounded players, no such strategy exists).

2.3 Examples

In this section, we present examples to illustrate the notion of renegotiation-safe profiles. As already mentioned, in normal-form games (bayesian or not), RS and NE, and thus also subgame perfect equilibria (SPE) coincide. We here illustrate the differences between these notions in the context of extensive-form games.

NE, but not SPE or RS The game in Fig. 1 has a *strict* Nash equilibrium (L, b) that relies on an "empty-threat"; this equilibrium is neither subgame perfect nor RS because the first player can renegotiate to (R, a). More generally, the game in Fig. 1 is a generic perfect information game and thus has a unique SPE. As we show in Corollary 9, in sequential games with unique SPE (and thus generic perfect information games), SPE and RS coincide.[2]

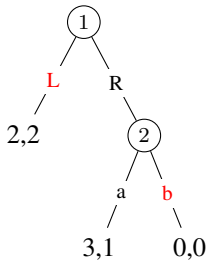

Figure 1: An empty threat

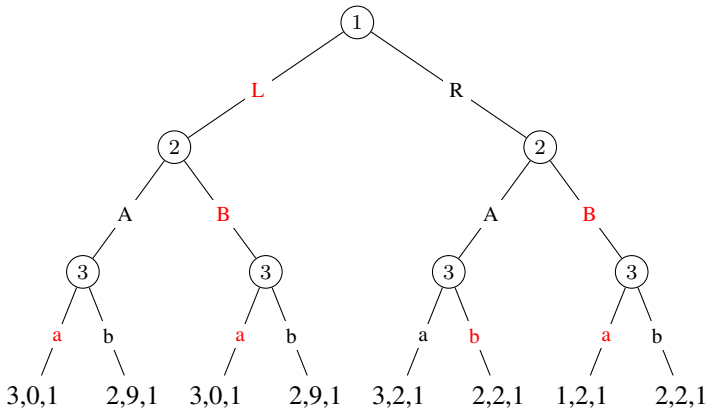

Figure 2: Game with an SPE that is not RS

SPE but not RS (non generic games with perfect information) In the game in Fig. 2, the profile $(L, (B, B), (a, a, b, a))$ highlighted in red is a SPE. The profile is not RS, however, because player 2 can renegotiate to $(L, (A, \cdot), (b, \cdot, \cdot, \cdot))$. Clearly this is a game of perfect information, but player 3's utilities are all 1 and so the game is not generic. This game has the RS $(R, (A, \cdot), (\cdot, \cdot, a, \cdot))$. More generally, as we show in Theorem 4, all sequential games have a RS strategies.

We now provide a more complicated example of a generic game of complete information where SPE and RS do not coincide. In fact, this game does not even have an RS.

SPE but not RS (generic games with complete, but imperfect, information) Consider game G depicted in Fig. 3. It is a finite normal-form game without a "maximal" NE—i.e., there is no NE that is preferred by both players. Let u_i^{max} be player i's expected utility in its best NE. Let G' be a cheap-talk extension of G where a simultaneous (cheap) move is followed by the players playing G. We claim that G'—a generic game without perfect information—does not have an RS strategy profile. Note that G' clearly has a NE and a SPE (and thus a sequential equilibrium), but these equilibria cannot be RS (since the game does not have any RS strategies).

[2] We thank Geir Asheim for asking whether SPE and RS coincide in generic perfect information games.

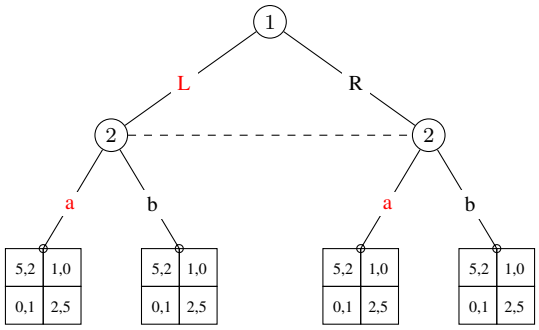

Figure 3: Game G'. In the first round, players move simultaneously and announce cheap-talk messages. In the second round, they play a normal form game.

Theorem 3. *Game G' is a finite 2-round 2-player extensive-form game of complete information without an RS strategy.*

Proof: Assume, for contradiction, that σ is RS for G'. We start by showing the following claim:

Claim 1. *There exists some player i such that $u_i(\sigma) < u_i^{max}$.*

Proof: We start by noting that RS strategies in G' induce a mixture of NE in G: Consider any strategy σ' in G'. By Lemma 1 we have that for every round 1 history h, $\sigma'(h)$ is RS in $G'(h) = G$. Since G is a normal-form game, by Fact 1, $\sigma'(h)$ is thus a NE, so σ' induces a mixture of NE in G.

Since σ is RS it thus also yields a utility profile that is a mixture of NE utility profile in G. Since G does not have a maximal NE, there thus exists some i such that $u_i(\sigma) < u_i^{max}$. □

But, if Claim 1 holds, player i has a RS renegotiation τ at round 1: in the first round, pick any action (say L); in the second round, play the NE that yield the best utility u_i^{max} for player i. This renegotiation clearly improves i's utility. We only need to argue that τ is RS at round 2. By Lemma 1, this amounts to showing that τ is RS in every subgame reached after the first round by $[\sigma, 1, \tau]_i$. By construction, all these subgames are the same. Since τ plays a NE in this subgame, we have by Fact 1 that this NE is also RS (at round 1), which concludes the Theorem. □

RS but not NE. The example in Fig. 4 shows a profile $(L, (a, a))$ which is a RS but is not even NE: player 1 would deviate to R given that player 2 plays a in both histories.

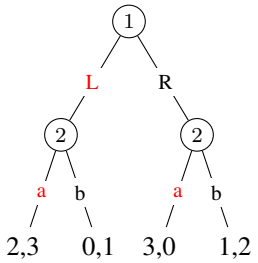

Figure 4: Examples that illustrate the RS concept

Of course, the strategy $(L, (a, b))$ leads to the same play as $(L, (a, a))$ and is both NE and RS. Indeed, as we show in Corollary 9, in sequential games, every RS σ can be *purified* into another RS σ' that generates exactly the same play as σ' but is also a NE—in fact, σ' is even a subgame perfect equilibrium. This purification, however, might not preserve the computational complexity of σ; in particular, the strategy σ' might be a lot more complex than σ.

To Conclude RS is incomparable to (i.e., neither strictly stronger than, nor strictly weaker than) both subgame perfection and Nash equilibrium.

3 Existence and Purification

Since for normal-form games, the RS concept is equivalent to NE, by Nash's theorem, every finite (Bayesian) game has a RS strategy. However, as demonstrated in the Thm. 3, this is not true for the case of extensive-form games.

In this section we show the existence of RS strategies for a natural class of games, namely sequential games (i.e., finite extensive-form games in which one player makes a publicly observable action per round, and in the last round, all players move simultaneously). Additionally, we present a "purification lemma" showing that for sequential games, any RS strategy σ can be purified into a startegy that is both RS and SPE, while yeilding the same distribution over outcomes; combining the two we thus have that every sequential game has a profile that is both RS and a SPE.

Note that games of perfect information (considered in [Fer95]) are special cases of sequential games. Another special-case of interest is "cheap-talk" extended normal-form games—where the players communicate, one after the other, over public-channels, and next play a normal-form game.

Our main existence theorem follows.

Theorem 4. *Every sequential game G has an RS strategy σ that is also SPE.*

Proof: We prove this theorem in two steps. We first show in Lemma 5 that G has an RS σ. In the second step, we show in Lemma 7 how to turn σ into a strategy profile that is both RS and SPE. □

3.1 Exitence of RS

Let $RS(G)$ denote the set of renegotiation safe profiles for G.

Lemma 5. *For every sequential game G, the set $RS(G)$ is non-empty and compact.*

Proof: Let G have q rounds. The proof follows by induction on the length of the game. We start by showing the base case when G is a normal-form game.

Claim 2. *For every finite normal-form game G, $RS(G)$ is non-empty and compact.*

Proof: By Nash's theorem, every finite normal-form game G has a NE and thus by Fact 1, $RS(G)$ is non-empty. Clearly the set $RS(G)$ is bounded as G is a subset of $([0,1]^{|S|})^n$ where S is the set of actions in G and n is the number of players. We turn to show that it is also closed. Assume not. That is, there exists a sequence of strategies $\langle \sigma^m \rangle_{m=1,2,..}$ converging to σ (in the sense that the probability placed by σ^m on a pure strategy profile converges to the probability placed by σ on that same strategy profile) such that $\sigma^i \in RS(G)$ but $\sigma \notin RS(G)$. This means there exists a player i and a strategy σ_i' such that $u_i(\sigma_i', \sigma_{-i}) - u_i(\sigma) > 0$, but for every m, $u_i(\sigma_i', \sigma_{-i}^m) - u_i(\sigma_m) \leq 0$. However, by continuity of expected utility $u_i(\sigma_i', \sigma_{-i}^m)$ converges to $u_i(\sigma_i', \sigma_{-i})$ and $u_i(\sigma_m)$ to $u_i(\sigma)$, which is a contradiction. We conclude that $RS(G)$ is closed and thus compact. □

We proceed to show the induction step.

Claim 3. *Consider some round q. Assume that $RS(G)$ is non-empty and compact for every r'-round sequential game G such that $0 \leq q' \leq q$. Then, for every $(q+1)$-round sequential game G, $RS(G)$ is non-empty and compact.*

Proof: Consider an $q+1$-round sequential game G. Let i be the player that moves in the first round. We start by showing that $RS(G)$ is non-empty, and then show compactness.

Non-emptiness. For each round 1 action a, pick a strategy $\sigma(a) \in RS(G(a))$ that maximizes i's expected utility in $G(a)$. Since by the induction hypothesis $RS(G(a))$ is compact for every a (since $G(a)$ is an q-round sequential game), and since expected utility is continuous, there exists at least one such element. Let σ denote the partial strategy for G (starting after round 1, and continuing until the end of the game) which on the history a plays $\sigma(a)$. Finally, pick the action s such that (s, σ) maximizes i's expected utility; such an action exists since the action space is finite.

We show that (s, σ) is RS at every round. Let us first show that it is RS at round 1. If not, there exists some strategy (τ, σ') (where τ is a distribution over actions for player i in round 1) which gives i higher utility than $a, \sigma)$ and is RS at every round $r > 1$. It is easy to see that, without loss of generality, we can assume that τ only assigns positive probability to a single action s'. Thus, there exists some strategy (s', σ') that is RS from round 2, such that

$$u_i(s', \sigma') > u_i(s, \sigma). \tag{1}$$

Since (s', σ') is RS from round 2, it follows by Lemma 1 that for the history s' (which is reached with probability 1 by (s', σ')), $\sigma'(s')$ is RS in $G(s')$. This means that

$$u_i(s', \sigma) \geq u_i(s', \sigma')$$

since $\sigma(s')$ is selected so as to maximize i's utility among RS strategies in $G(s')$. Furthermore, since s is also selected so as to maximize i's utility (given the continuation σ), we have that,

$$u_i(s, \sigma) \geq u_i(s', \sigma) \geq u_i(s', \sigma')$$

which contradicts equation 1.

We move on to show that (s, σ) is RS from round 2. This directly follows by Lemma 1 since $\sigma(s)$ was selected from $RS(G(s)))$ and then only length 1 history (s, σ) assigns positive probability to is s.

We proceed to show that $RS(G)$ is compact.

Compactness. It easily follows as in Claim 2 that that $RS(G)$ is bounded since the space of pure behavioral strategies is bounded. We turn to show that it is also closed. Assume not. That is, there exists sequence of strategies $\langle \sigma^m \rangle_{m=1,2,..}$ converging to σ such that $\sigma^j \in RS(G)$ but $\sigma \notin RS(G)$.

Since $\sigma^j \in RS(G)$ it follows that for every j, player i's utility of σ^j in G is the same, and thus by continuity of expected utility (and compactness of the strategy space), the utility of σ in G is the same, which means that σ is RS at round 1.

It only remains to show that σ is RS from round 2. Assume not. By Lemma 2 there thus exists some non-empty history h such that $\sigma(h)$ is not RS in game $G(h)$. Since σ assigns positive probability to h, we can assume without loss of generality that, for *all* j, σ^j assigns positive probability to h as well (or else σ^j cannot converge to σ). Since σ^j is RS and assigns positive probability to the history h, it follows by Lemma 2 that $\sigma_j(h)$ is RS in $G(h)$. Finally, by compactness of $RS(G(h))$, which follows from the induction hypothesis (since h is non-empty and thus $G(h)$ is a q-round game), it follows that $\sigma(h) \in RS(G(h))$, which is a contradiction.[3] $\square$

These two claims establish that $RS(G)$ is non-empty and compact, and therefore contains some profile σ. $\square$

[3]Note that this last step relies on the fact that we consider games of complete information (i.e., non-bayesian games). If we had considered a bayesian game, we could only conclude that $\sigma_j(h)$ is RS in $G_{\sigma_j}(h)$, but since $G_{\sigma_j}(h)$ and $G_\sigma(h)$ might be different games (if G is bayesian), the above argument fails.

3.2 Purification

We now use the existence lemma to prove the purification lemma previously mentioned. We first show that in sequential games G, any RS can be "purified" into a strategy profile σ' that generates the same play as σ, but σ' is RS in $G(h)$ for *any* history h (even those that are not reached by σ; recall that in constrast, by Lemma 2, σ is only RS in $G(h)$ for histories h that are reached by σ).

Lemma 6. *Let G be a sequential game and let σ be a RS strategy profile in G. Then there exists a strategy profile σ' that leads to same distribution over outcomes in G such that σ' is RS in $G(h)$ for every non-terminal history h in G.*

Proof: Let σ' be equal to σ on all histories that are reached by σ. Now iteratively proceed as follows until σ' is fully defined: for all histories h on which σ' currently is undefined (at first this is just the histories h that are not reached by σ), let $\hat{h}$ be the shortest prefix of h for which σ' is undefined, pick a RS strategy τ in the game $G(\hat{h})$ (such a RS strategy exists by Theorem 4 and define σ' to play τ on $\hat{h}$ and every history that is reached by τ from $\hat{h}$. Clearly σ' is still RS (since we have only modified σ on histories that are never reached by σ). In fact, by construction σ' is trivially RS in $G(\hat{h})$ for every history $\hat{h}$ used in the purification process. Finally, by applying Lemma 2, we have that σ' is RS in $G(h)$ also for all the histories h reached by τ. This concludes that σ' is RS in $G(h)$ for all histories h. □

Purified strategies σ' are also SPE:

Lemma 7. *Let G be a sequential game and let σ a strategy profile such that σ is RS in $G(h)$ for every non-terminal history h in G. Then σ is also a subgame perfect equilibrium.*

Proof: Assume not; that is, there exists a history h and a deviation σ_i' for player i that improves i's utility in $G(h)$. By the one-step deviation principle, we can assume without loss of generality that σ_i' is identical to σ_i on every history except for h. As we now argue, this means that σ cannot be RS at round 1 in $G(h)$, which is a contradiction. Consider the renegotiation to (σ_i', σ_{-i}) in round 1. By definition, it increases player i's utility. Furthermore, since σ_i' is identical to σ on every history except for h and since σ is RS at every history, by Lemma 1 we have that (σ_i', σ_{-i}) is RS from round 2. □

Combining the two lemmas we directly have the following corollary:

Corollary 8. *Let G be a sequential game and let σ be a RS strategy profile in G. Then there exists a strategy profile σ' that leads to same distribution over outcomes in G such that σ' is both RS and SPE.*

As a simple consequence, we have:

Corollary 9. *If G is a sequential games with a unique SPE σ, then σ is also RS. Thus, in generic games of perfect information, every SPE σ is also RS.*

Proof: By Theorem 4, G has an RS σ. Purify σ to be an SPE by Lemma 7. Since there is only one SPE in G, this purification of σ must correspond to τ, and therefore τ must also be RS. (Finally, it is well known that generic games of perfect information have a unique SPE.) □

4 Renegotiation-Safe Secret Sharing

A (t, n) secret-sharing scheme consists of two efficient (probabilistic) algorithms SHARE and UNSHARE. The SHARE(δ) method produces shares $(s_1, \ldots, s_n)$ such that any subset of less than t shares reveals no information about δ. The UNSHARE algorithm outputs either a string or the special failure symbol $\perp$ and does the opposite: for any secret δ, any sharing $S = (s_1, \ldots, s_n) \leftarrow$ SHARE(δ), and any set of t valid shares $\{s_1', \ldots, s_t'\} \subseteq S$, it holds that UNSHARE$(s_1', \ldots, s_t') = \delta$.

In an *authenticated* variant of secret sharing, SHARE outputs an extended share $\hat{s}_i = (s_i, y_i, v_i)$ for each player such that player i can use verification information v_i against player j's authentication information y_j to determine if player j's announced share s_j is valid. In particular, there exists a VER function so that if $(\hat{s}_1, \ldots, \hat{s}_n) \leftarrow \text{SHARE}(\delta, \epsilon)$, then for any j, $\text{VER}(v_i, (s_j, y_j)) = 1$. In addition to these syntactic requirements of a secret sharing scheme, the following definition captures the security goals:

Definition 4 (Authenticated Secret Sharing Security)**.** We say that a (t, n)-authenticated secret sharing scheme $(\text{SHARE}, \text{UNSHARE}, \text{VER})$ is (t, n, ϵ)-secure if for any two secrets $\delta_0, \delta_1 \in \Delta$, and any subset $X \subset [1, n]$ of size $t - 1$, the following two distributions are identical:

$$\{(\hat{s}_1, \ldots, \hat{s}_n) \leftarrow \text{SHARE}(\delta_0) : \{\hat{s}_i\}_{i \in X}\}$$
$$\{(\hat{s}_1, \ldots, \hat{s}_n) \leftarrow \text{SHARE}(\delta_1) : \{\hat{s}_i\}_{i \in X}\}$$

and for any j and any secret $\delta \in \Delta$,

$$\Pr[(\hat{s}_1, \ldots, \hat{s}_n) \leftarrow \text{SHARE}(\delta); (s', y') \leftarrow A(\hat{s}_{-j}) : \text{VER}(v_j, (s', y')) = 1 \wedge s' \notin \{s_i\}] < \epsilon$$

Notice the definition assumes that the dealing step is performed honestly.

Example The (t, n, ϵ)-secure secret sharing protocol due to Shamir works as follows. The $\text{SHARE}(\delta)$ function samples a random polynomial p of degree $t - 1$ such that $p(0) = \delta$. Set the share $s_i = p(i)$ for $i = 1, \ldots, n$. To UNSHARE, a group of t players interpolate a polynomial p' (using Lagrange) through their t points and then compute $p'(0)$. The scheme can be augmented with a Message Authentication Code (MAC) so that each player i, in addition to receiving s_i, also receives authentication information for its own shares and MACs for each of the other players' shares as follows: pick a K-bit prime number p such that $2^{-K} < \epsilon$. Generate player i's MAC for player j's share by choosing a pair of numbers $a_{i,j}, b_{i,j} \in \mathbb{Z}_p$ and compute $y_{i,j} = s_j \cdot a_{i,j} + b_{i,j} \bmod p$. Player i receives the pair $v_{i,j} = (a_{i,j}, b_{i,j})$ which describes a line, and player j receives $y_{i,j}$, which along with s_j, represents a point on the line. Thus, a share generated by SHARE contains $\hat{s}_j = ((s'_j, y_{1,j}, \ldots, y_{n,j}), (v_{1,j}, \ldots, v_{n,j})$. For convenience, we sometimes write this as $s_j = (s'_j, y_j, v_j)$.

Notice that j cannot generate another point on the line with probability any better than ϵ, and player i learns nothing about j's point since the pair $(a_{i,j}, b_{i,j})$ are just random numbers. With these extra authentication values, $\text{VER}(v_i, (s_j, y_{i,j}))$ can test whether a putative share s_j was a valid by checking that $(s_j, y_{i,j})$ is a point on the line defined by $v_{i,j} = (a_{i,j}, b_{i,j})$.

We thus have:

Theorem 10 (Shamir)**.** *For any $n \geq 2$, $t \leq n$, and $\epsilon > 0$, there exists a (t, n, ϵ)-authenticated secret sharing scheme.*

4.1 The Reconstruction Process

The question of existence and efficiency of secret-sharing schemes is a purely cryptographic one that has been well studied in the cryptographic literature. We here focus on a separe question; namely, how the players can jointly reconstruct the secret.

We envision the algorithms of a secret-sharing scheme to be used in a setting in which an honest dealer has prepared an unguessable secret δ, uses the SHARE algorithm to produce extended shares $(\hat{s}_1, \ldots, \hat{s}_n)$, distributes the shares to n parties, and then disappears. At a later point, a subset of these players desire to reconstruct δ. If they can cooperate in order to disseminate t of their shares among each other, then they can use the algorithm UNSHARE to accomplish the goal. Of course, if the players have a trusted mediator, then they can cooperate. However, the principal game theoretic question (first studied in [HT04]; see Appendix A for a brief overview of related works) is whether there exists a "rational" *cheap-talk process* by which this

cooperation can be accomplished—especially in the case when players prefer that fewer people learn the secret, i.e., when utilities are strictly competitive. We use a cheap-talk process because we assume that a player's communication does not affect their utility; rather, only the types and the final outcome of who learns the secret determines utility.

The straightforward cheap-talk process is to have each player broadcast their share one after the other. If a player sends an invalid share before round t, then the process ends with failure. However, as argued by Halpern and Teague [HT04], the obvious problem is that as soon as $t-1$ shares have been broadcast, the remaining players who have yet to send their shares now learn the secret. They no longer have any reason to broadcast their own shares since utilities are strictly competitive.

Despite this shortcoming, this protocol is a Nash equilibrium when $t \leq n-1$ because as long as the remaining $n-1$ players play the Nash strategy, it is still a best response to broadcast. (In fact, we conjecture that a slight modification of this weak protocol is also a sequential equilibrium.) As we now proceed to argue, the weakness of this specific process is intrinsic in the sense that no sequential process for secret share reconstruction can be renegotiation-safe.

4.2 Guessing Game

We generalize the process of secret-share reconstruction to the following type of "guessing games". We consider a sequential bayesian game where the players attempt to guess the type of nature. The utilities are "strictly competitive"; namely, players want to recover the correct secret, and if they do, they prefer that fewer players recover it. Given the type profile $(\theta_0, \theta_1, \ldots, \theta_n)$, we say that player i recovers (or gets) the secret at the terminal history o if in o, the final action by i is θ_0; let $\mathsf{R}(o)$ denote the set of players that retrieve the value at o.

Definition 5 (Guessing game)**.** Π is called a (n, r, μ, u, P)-*guessing game* if Π is an $r+1$-round, n-player finite sequential bayesian game with player function P, type distribution μ and utility functions such that for player i, $u_i(o) = g_i(b, j)$ where the bit $b = 1$ iff $i \in \mathsf{R}(o)$ and $j = |\mathsf{R}(o) - \{i\}|$ and g_i is a linear function satisfying the following two properties:

1. (learning is preferred) : $g_i(1, j) > g_i(0, j)$ for any j.
2. (scarcity is strictly preferred) $g_i(b, j) > g_i(b, j')$ for any $b \in \{0, 1\}$ and $j < j'$,

Π is called an r-round guessing game if it is a (n, r, μ, u, P)-guessing game for some n, μ, u, P.

Remark 7. Note that the fact that g_i is linear means that player i's utility under σ can be written as

$$u_i(\sigma) = c_1 \cdot \Pr_{\sigma}[i \in \mathsf{R}(o)] - c_2 \cdot \mathbb{E}_{\sigma}[|\mathsf{R}(o) - \{i\}|]$$

for some positive constants c_1, c_2.

Note that an r-round guessing games is an r-round cheap-talk extension of a one-shot game: the actions in the first r rounds are used only for communication (we refer to these rounds as the communication rounds), and utility is only defined based on the actions the player choose in the $r+1$'st (simultaneous) move.

Also, note that if the players do not communicate at all, then the best strategy for player i is to simply guess the sought value (independently of what everyone else is doing). For player i, let p_i denote the highest success probability i can have in guessing the sought value. We will be interested in guessing games where the players prefer jointly retrieving the sought value to trying to guess it (even if no one else gets it). More formally, we say that *collaboration is preferred* if for all player i, $g_i(1, n-1) > \mathbb{E}(g_i(B, 0))$ where B

is random variable that is 1 with probability p_i (and 0 otherwise). [4] Note that since guessing games are cheap-talk games, only μ and u determine whether collaboration is preferred in the game.

4.3 Impossibility of RS-Collaboration in Guessing Games

We show that collaboration cannot be acheived by RS profiles in guessing games. We will later rely on this results to construct a ϵ-RS rational secret sharing scheme assuming two non-negotiating players.

Given an r-round guessing game Π, let Π^x denote Π modified to only keep the first $x \leq r$ communication rounds of Π (but leaving everything else the same). For instance, $\Pi = \Pi^r$, Π^{r-1} is identical to Π but without the last communication round, and Π^0 is simply a one-shot bayesian game without communication.

Theorem 11. *If Π is an r-round guessing game and σ is an RS profile for Π, then Π^{r-1} has an RS profile with the same expected utility as σ.*

Before proceeding to the proof of Theorem 11, let us first prove the following useful lemma. Given an r-round guessing game Π and a strategy σ for Π, let σ^{r-1} denote σ modified into a strategy for Π^{r-1}: the last player j to move in Π simply does not send its last message and in the final round all players make the best possible guess for the secret assuming that the history of messages was generated using σ (up until round $r-1$). Since the strategy space is compact, such a best guessing strategy exists for every player.

Lemma 12. *Let Π be an r-round guessing game and let σ be a profile that is RS at round $r+1$ in Π. Then, $u'_j(\sigma^{r-1}) \geq u_j(\sigma)$, where u and u' are the utility functions in Π and Π^{r-1} respectively and j is the player moving in round r. If furthermore σ is RS at round r, then for* every *player i, $u'_i(\sigma^{r-1}) = u_i(\sigma)$.*

Proof: We start by showing part 1. Assume that σ is RS at round $r+1$. We have that for every player $s \neq j$, the probability that s outputs the secret in σ' is no greater than the probability that it outputs it in σ, or else s should simply renegotiate to $[\sigma, r+1, \sigma^{r-1}]_s$ (i.e., play σ as before, but simply ignore the last message from j)—this increases s probability of retreiving the secret, and the expected number of other player that do stays the same, and thus, by linearity of u_i (see Remark 7), s will improve its utility; furthermore, the renegotiation is trivially stable. Thus, by the linearity of expectations, the expected number of players excluding j that retrieve the secret cannot go up when moving from σ to σ^{r-1}, and trivially, the probability that j retreives the secret is the same in σ and σ^{r-1}. By linearity of u_j, we thus conclude that $u'_j(\sigma^{r-1}) \geq u_j(\sigma)$.

We move on to show part 2. Assume further that σ is RS also at round r. Then we also have that $u'_j(\sigma^{r-1}) = u_j(\sigma)$ since otherwise j would prefer to renegotiate to σ^{r-1} in round r (which by construcion is a stable renegotiation). Applying linearity of utilities again, we have that for each player $s \neq j$, the probability that s outputs the secret in σ^{r-1} is no smaller than the probability that it outputs it in σ, or else the expected number of player exlcuding j that retreive the secret goes down and thus $u'_j(\sigma^{r-1}) > u_j(\sigma)$. Thus, for each player i we have that the probability that it retreives the secret is the same in σ and σ^{r-1}, and thus (by another application of linearity of utilities), $u_i(\sigma) = u'_i(\sigma^{r-1})$. $\square$

We now proceed to the proof of of Thm. 11.

Proof: (*of Thm. 11*) Let j be the player speaking in the last communication round in Π, $\Pi' = \Pi^{r-1}$ (i.e., Π without the last communication round), u, u' be the utility functions for Π, Π' respectively, and let $\sigma' = \sigma^{r-1}$. We show that σ' is an RS for Π' with the same expected utility as σ. By Lemma 12 we directly have that σ and σ' have the same utility. We proceed to show that σ' is also RS.

Suppose for contradiction that σ' is not RS in Π'. This means there exists a stable renegotiation ρ' for some player s at round ℓ in the shorter game Π' so that $u'_s([\sigma', \ell, \rho']_s) > u'_s(\sigma')$. Let ρ be a renegotiation for

[4] An even weaker condition potentially suffices for our results: $g_i(1, n-1)$ is only required to be greater than the best utility i can get in any NE where there is no communication among the players; this value is smaller than $E(g_i(B, 0)$ since the utility of i goes down when the other players also manage to guess the secret.

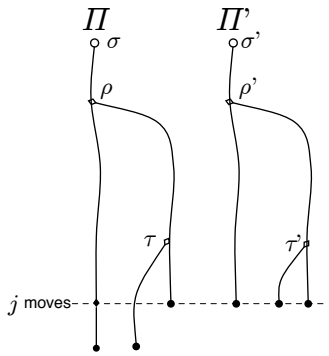

Figure 5: Diagram of the strategies used in the impossibility argument

the long game Π that mimics ρ' and in addition instructs player j not to send the last message. Since ρ' and ρ induce the same actions in the games Π' and Π respectively, we have that $u_s([\sigma, \ell, \rho]_s) > u'_s(\sigma') = u_s(\sigma)$. Thus, switching to ρ is a profitable renegotiation for s in Π. But because σ is RS, the renegotiation ρ cannot be stable. There must thus exist a stable renegotiation τ from ρ at a round $\ell' > \ell$ by some player t (that need only result in better utility for player t). If τ_j instructs j not to send the last message, then τ also corresponds to a stable renegotiation τ' with respect to ρ' in the short protocol, which contradicts that ρ' was a stable renegotiation in the short protocol. We conclude the proof by showing that, without loss of generality, τ_j in fact does instructs j to stay silent in the last communication round.

Claim 4. *Without loss of generality, we can define τ so that j stays silent in round r.*

Proof: Suppose that τ_j instructs j to send a message in the last round. First note that this means that $\ell' \leq r$ (since ρ instructs j to stay silent). Redefine $\tau = \hat{\tau}$ so that j no longer sends the last message and finally all players make the best possible guess for the secret assuming that the history has been generated by running (σ, ρ, τ); again, since the strategy space is compact, such a best guessing strategy exists for every player. We now show that t's utility can only increase by switching to $\hat{\tau}$ (and thus $\hat{\tau}$ is still a profitable renegotiation); furthermore $\hat{\tau}$ is RS for every round $l > \ell''$. Note that since after round r, $\hat{\tau}$ outputs the best possible guess for the secret (by construction), $\hat{\tau}$ is RS at round $r + 1$. Now, consider the following two cases.

Case 1: $\ell' = r$. In this case $t = j$. Since (σ, ρ, τ) is RS at round $r + 1$ (since τ is a stable renegotiation), it directly follows from part 1 of Lemma 12 that j's utility can only increase. And since $\hat{\tau}$ is RS at round $r + 1$, $\hat{\tau}$ is stable.

Case 2: $\ell' < r$. In this case, (σ, ρ, τ) is RS at both round r and round $r + 1$, so it directly follows from Lemma 12 that for every player i (and thus also for player t), its utility remains unchanged. It only remains to show that $\hat{\tau}$ is RS at every round $l > \ell'$. We already know that it is RS at round $r + 1$. So if there is a renegotiation ψ for some player q, it must happen at a round $l \leq r$; but since τ and $\hat{\tau}$ are identical before round r, and since they yield the same utility for all players (and in particular for q), we have that ψ would be a renegotiation also τ.

□ □

Corollary 13. *If Π is an r-round guessing game and σ is an RS profile for Π, then Π^0 has a Nash equilibrium with the same expected utility as σ.*

Proof: Apply Thm. 11 to strategy profile σ until there are 0 communication moves. By Fact 1 RS and Nash coincide for bayesian normal-form game; the corollary follows. □

Notice the proof does not rely on whether broadcast channels or point-to-point channels are used.

Remark 8. The argument relies on the notion of RS and cannot be applied to the Nash solution concept. For example, when we construct the profile σ', it may involve deviations by more than one person (i.e. player who guess and player j who is instructed not to send the last message) whereas Nash equilibria only study single-player deviations. Indeed, there exists a Nash equilibrium for the secret sharing reconstruction. (As mentioned, for any $n \geq 3$ and $t < n$, the simple 1-round broadcast protocol suffices: Given that that $n - 1 \geq t$ participants are honest, one deviation does not change the outcome.)

5 Renegotiation-Safe Secret Sharing with 2 Non-negotiating Players

In this section, we present an authenticated secret-sharing scheme that also has a reconstruction process that is ϵ-NE and also ϵ-RS when *two* out of n players do not renegotiate. The idea of considering non-negotiating players is similar to the idea of consider honest players in the cryptographic literature (going back to the work of Goldreich, Micali and Wigderson [**?**]). More recently, the BAR model of Aiyer et al [AAC$^+$05] considers a mixture of rational and honest player; even more recently, Ong et al in [OPRV09] rely on this idea in the context of rational secret sharing, showing a protocol that satisfies a notion of trembling hand perfection when $\omega(\log n)$ players are honest.

5.1 Modeling Non-negotiating players

To model non-negotiating (or "honest") players, we consider an extension of the standard bayesian game model. We fix the honest players strategy σ (no matter what strategy the player actually "chooses"). To formalize this, we rely on the idea of *interpreted strategies* from Halpern and Pass [HP10b]. More precisely, we extend our model of bayesian games (see Section 2.1) as follows:

- For each player i, we add an *interpretation function* I_i to the game; $I_i : S \to S$ is a function that maps strategies to strategies. When evaluating the expected utility of a strategy profile σ, the actions of player i given type i) are computed using the strategy $I_i(\sigma_i)$ (instead of just σ_i).

We call such games *extended* bayesian games. We will focus on two types of interpretation function:

- "Normal" (i.e., rational) players' interpretation function is simply the identity function $I(\sigma) = \sigma$.
- "Honest" players' interepretation function is χ^i_τ for some strategy τ, where $\forall \sigma \in S$, $\chi_\tau(\sigma) = \tau$.

We call an extended bayesian game τ-*honest extended* if for all player i, their interpretation functions is either identity or χ_{τ_i}. We call an extended bayesian game (k, τ)-*honest extended* if it is τ-honest extended, and at least k players i use the honest interpretation function χ_{τ_i}. Finally, we may also extend the notion of guessing games to (k, τ)-honest extended guessing games.

5.2 Protocol Π

We show that there exists a secret-sharing scheme and a reconstruction protocol σ such that for any distribution Δ over secrets for which the players prefer the secret to be reconstructed to simply guessing it (i.e., the secret has sufficiently high min-entropy), then assuming that at least 2 players are non-negotiating, σ is both a ϵ-NE and ϵ-RS, where the secret is *always* reconstructed.

Idea The idea of our protocol is to run many instances of an authenticated secret sharing reconstruction protocol (where the players simply reveal their shares one by one) in parallel such that each instance uses independently generated shares. In some instance, we guarantee that the two honest players are the last to

broadcast. Thus, as long as $t-2$ shares from that instance are broadcast, the honest players will ensure that the remaining two are broadcast so that all players learn the secret.

The schedule of player moves is chosen so that before $t-1$ shares have been broadcast in one instance of the reconstruction (and some player can now retrieve the secret), at least $t-2$ shares have been broadcast in all other instances. We call the point before $t-1$ shares are first broadcast the *critical point*.

Roughly speaking, this protocol is RS because a renegotiation requires some player to deviate from the protocol either before the critical point, or after it. If before, the two honest players stop talking, but no player has more than $t-1$ shares. Thus, the remaining players find themselves in a guessing game whose only RS (by Thm. 11) has poor utilities. Thus, such a renegotiation is not profitable. If the deviation occurs after the critical round, at least $t-2$ shares have been broadcast in all instances. For one of those instances, the last two players to broadcast are honest and ensure that everyone retreives the secret. Thus, such a renegotiation does not improve anyone's utility.

The protocol is only an ϵ-RS and not an RS because the authentication part of the secret sharing scheme can be broken with a small probability. Thus, a renegotiation may try to fake a share (after the critical round); if successful, this renegotiation may cause the other players to fail in reconstructing the secret. By the security of the authenticated secret sharing scheme, this even occurs with a very small probability. (In fact, the strategy of *optimally*[5] trying to fake the share is a *true* RS strategy in this game, assuming the honest players still honestly reveal their shares, as before.)

The details. To formalize the above discussion, we use the following notation. If μ is a distribution over $(\theta_1,\ldots,\theta_n)$, then the notation $f(\mu)$ represents the distribution resulting from first sampling an n-tuple $(\theta_1,\ldots,\theta_n)\leftarrow\mu$, and then returning $(\theta'_1,\ldots,\theta'_n)\leftarrow f(\theta_1,\ldots,\theta_n)$.

Definition 6. Let $\Pi=(\text{SHARE},\text{UNSHARE},\text{VER})$ be a secure (t,n,ϵ)-authenticated secret-sharing scheme, σ a strategy profile, and let Δ,u be such that u is the utility function for a guessing game where (a) the type distribution samples the type of nature according to Δ, and all other types are empty, and (b) collaboration is preferred; let $\mu=\text{SHARE}(\Delta)$.

We say that σ is an *r-round k-honest ϵ-rational reconstruction protocol* for scheme Π and player function P if: for any (n,r,μ,u,P)-guessing game G where the action space is compatible with σ, and any (k,σ)-honest extended (n,r,μ,u,P)-guessing game $\hat{G}$ where the action space is compatible with σ, it holds that:

1. σ leads to all players always retrieving the secret in G and $\hat{G}$.
2. σ is an ϵ-NE in G.
3. σ is ϵ-RS in $\hat{G}$.

Consider the scheme $\Pi(\pi,n)=(\text{SHARE}^{\pi,n},\text{UNSHARE}^{\pi,n},\text{VER}^{\pi,n})$, the player function P_n, and the strategy $\sigma^{\pi,n}$, described in Fig 6 and Section 5.2.1 respectively. We have the following proposition.

Proposition 1. *Let π be a $(t,n,\epsilon/2n^2)$-secure authenticated secret sharing scheme. Then, $\Pi(\pi,n)$ is a (t,n,ϵ)-secure authenticated secret sharing scheme such that $\sigma^{\pi,n}$ is an ℓ-round 2-honest ϵ-rational reconstruction protocol for Π and player function P_n where $\ell=n(n-1)/2$.*

Note that this proposition applies to any threshold t including $t=n$ (assuming the underlying secret sharing scheme is secure for (t,n)); recall that, in contrast, the simple broadcast protocol discussed in Section 4.1 is not a NE for $t=n$.

Finally, combining Proposition 1 with Theorem 10 implies:

[5]Note that if we use the particular authenticated secret sharing scheme from Thm. 10, such an optimal strategy is easy to implement—a random guess for the authentication information is optimal.

$\Pi(\pi)$ uses a $(t, n, \epsilon/2n^2)$ authenticated secret sharing scheme $\pi = (\text{SHARE}, \text{UNSHARE}, \text{VER})$.

SHARE$'(\delta)$: To produce shares for a t-out-of-n scheme of a secret $\delta \in \{0,1\}^k$ for $P_1, \ldots, P_n$

1. Set $\ell = n(n-1)/2$.
2. For each row $k = 1, \ldots, \ell$, repeat the following
   (a) Generate a random t-sharing $(\hat{s}_{k,1}, \ldots, \hat{s}_{k,n}) \leftarrow \text{SHARE}(\delta)$ of the secret δ using an authenticated sharing scheme where each subshare $\hat{s}_{k,i} = (s'_{k,i}, y_{k,i}, v_{k,i})$. (See section 4.)
   (b) Add subshare $\hat{s}_{k,i}$ to player i's final share.
3. Player i's final share $\hat{s}_i$ consists of $n(n-1)/2$ subshares, each of which is an independently generated share of the secret δ. I.e., share $\hat{s}_i = (\hat{s}_{1,i}, \ldots, \hat{s}_{\ell,i})$.

UNSHARE$'(\hat{s}_1, \ldots, \hat{s}_t)$: Output $\text{UNSHARE}(\hat{s}_{x,1}, \ldots, \hat{s}_{x,t})$ for the smallest index x such that all t subshares for row x verify.

VER$'(\hat{s}_i, (s', y'))$: Compute $\bigwedge_{x=1}^{\ell} \text{VER}(v_{x,i}, (s'_x, y'_x))$ to verify all of the subshares within s'.

Figure 6: Secret Sharing scheme Π

Theorem 14. *For every $n > 2, t \leq n, \epsilon > 0$, there exists a (t, n, ϵ)-secure authenticated secret sharing scheme Π and a 2-honest ϵ_1-rational reconstruction protocol for Π.*

Let us first argue that Π is a secure scheme.

Lemma 15. *$\Pi(\pi) = (\text{SHARE}', \text{UNSHARE}', \text{VER}')$ is a (t, n, ϵ)-secure authenticated secret sharing scheme when π is $(t, n, \epsilon/2n^2)$-secure.*

Proof: The lemma follows from (a) the fact that the n-fold direct products of two identical distributions results in two distributions that are also identically distributed and (b) the union bound which establishes an upper bound on the verification error VER$'$ based on the verification error of VER. $\square$

5.2.1 Reconstruction process σ

Player function P_n We first describe the player function P_n that maps a history to a player (i.e. that defines the order in which the players broadcast messages) as a matrix O of size $n \times \ell$ where $\ell = n(n-1)/2$. Each row of the schedule is a permutation of the n players. The rows of O are chosen such that the last two columns of O contain every pair (i, j) where $i < j$ and $i, j \in [1, n]$. In other words, for every pair of players, there is some row in O for which the pair appears at the end of the row. For example, when $n = 4$, one choice for O could be

$$\begin{array}{cccc} 3 & 4 & 1 & 2 \\ 2 & 4 & 1 & 3 \\ 2 & 3 & 1 & 4 \\ 1 & 4 & 2 & 3 \\ 1 & 3 & 2 & 4 \\ 1 & 2 & 3 & 4 \end{array}$$

Observe for any pair (i, j) with $i < j$ and $i, j \in [1, 4]$, there is some row of O which ends with i, j.

Function P_n is defined by the columns of schedule O. First, the player identified by $O_{1,1}$ broadcasts, then the player identified by $O_{2,1}$, and so forth until $O_{\ell,1}$. At this point, we say that the first column has

completed. Next, player $O_{2,1}, O_{2,2}, \ldots, O_{2,\ell}$ broadcast their messages, then $O_{3,1}, \ldots$, and so forth until the entire schedule O has been executed in this manner.

Critical point The critical point in such a schedule is the point before which some player is the first to broadcast in column $t-1$.

Reconstruction strategy $\sigma_i(\hat{s}_i)$: The strategy σ_i also makes use of the scheme $\pi = (\text{SHARE}, \text{UNSHARE}, \text{VER})$ used in the construction of Π. The strategy is described in two phases:

1. (Before critical point, e.g. first $t-2$ columns) At step $O_{j,k}$ for $k \in [1, t-2]$: Player $p = O_{j,k}$ runs VER (for the scheme π) on every share that has been broadcast using the verification information v_p in p' share. (Interpret an abort or unreceived message as a 0.) If any player has broadcast a message that does not verify, then halt (and do not continue to Phase II). Otherwise, broadcast the authenticated share $(s_{j,p}, y_{j,p})$ given by the dealer to player p for row j.
2. (After critical point) At step $O_{j,k}$ for $k \in [t-1, n]$: Player $p = O_{j,k}$ determines if all messages before the critical point verified using the VER algorithm from π. If so, then p broadcasts the share $(s_{j,p}, y_{j,p})$ given by the dealer to p for row j. Otherwise, halt.
3. (End) After all parties have broadcast, find the first row in which t verified shares have been sent. If such a row exists, reconstruct the secret using these shares and UNSHARE and output the resulting value.

Proof: (of Proposition 1) Observe that σ leads to all players learning the secret. We separate the main argument into two parts. We first show that σ is an ϵ-RS and then argue that it is an ϵ-NE.

Claim 5. *Profile σ is an ϵ-RS for any $(2, \sigma)$-honest extended (n, r, μ, u, P_n)-guessing game where the action space is compatible with σ.*

Proof: Assume that there exists a renegotiation σ' by player j at round r which increases j's utility by at least ϵ, i.e., $u_j([\sigma, r, \sigma']_j) > u_j(\sigma) + \epsilon$. We distinguish between two cases:

Case 1: r is after the critical point In this case, the two honest players will always broadcast their share, which means that t correct shares are broadcast in some instance. If j gains by at least ϵ, there must exist some player j' who does not retrieve the secret with probability at least ϵ since utilities are in $[0, 1]$. But, then j' could improve its utility by running the reconstruction algorithm—by the security of the authenticated secret-sharing scheme, this guarantees that j' gets the secret with probability at least $1 - \ell\epsilon$ (the ℓn factor comes from the fact that there are at most ℓ shares whose authenticity can be "faked").

Case 2: r is before the critical point In this case, the two honest players stop communicating after round r. Consider a new game G' which has r less communication rounds than G, has the same utility function as G, but where the type distribution is as follows: sample types θ just as in G, then concatenate an r-round history h to all players types, where h is sampled as the public history obtained by running $\sigma(\theta)$ for $r-1$ rounds, and finally generating the round r message by running σ'. Since $[\sigma, r, \sigma']_j$ is RS from round $r+1$, we directly have that σ' is RS in G'. By Thm. 11, we thus have that $u_j[\sigma, r, \sigma']_j$ is bounded by the best NE in G'^0 (i.e., the game G' but without any communication). Since utilities are strictly competitive, we can upperbound this utility by simply considering j's probability of guessing the secret and assuming no one else gets it. In other words, $u_j([\sigma, r, \sigma']_j) \leq \mathbb{E}(g_j(B, 0))$, where g_j is the linear function determining the utility u_j of player j, and B is a random variable that is 1 with the probability that player j can recover the secret. Now, since the only information j has received

in its type is its shares, and at most $t-2$ other shares, it follows by the security of the secret-sharing scheme, that $\Pr[B=1] \leq \mathsf{guess_j}$, where $\mathsf{guess_j}$ is j's a-priori probability of guessing the secret (without knowledge of any shares. Finally, since Δ, u are such that collaboration is preferred, and since utilities are in $[0,1]$, we have that

$$u_j([\sigma, r, \sigma']_j) \leq g(1, n-1) = u_j(\sigma).$$

□

Claim 6. *Profile σ is an ϵ-NE for any (n, r, μ, u, P_n)-guessing game where the action space is compatible with σ.*

Proof: Suppose there exists a strategy σ'_i for some player i such that $u_i(\sigma'_i, \sigma_{-i}) > u_i(\sigma) + \epsilon$. Towards reaching a contradiction, we consider a sequence of mental experiments, where we transform σ' into a strategy that is easier to analyze. In the following, let $\epsilon_1 = \epsilon/2n^2$.

Explicit deviation. We start by transforming σ'_i into a σ''_i which is identical to σ'_i but it explicitly signals any deviation from σ_i. More precisely, σ''_i is the same as σ'_i except that for any history h such that $\sigma'_i(h)$ and $\sigma_i(h)$ differ, $\sigma''_i(h) = \perp \cdot \sigma'_i(h)$ where $\perp$ is a special symbol that is not part of the alphabet in the game.[6] That is, σ''_i prepends the special symbol $\perp$ to any message sent that was not prescribed by σ_i. When players running σ_{-i} encounter a message prepended with $\perp$, they interpret this as an abort.

Claim 7. $u_i(\sigma''_i, \sigma_{-i}) \geq u_i(\sigma) + \epsilon - n^2\epsilon_1$.

Proof: The only difference in the actions taken between the two profiles occurs when some player $j \neq i$ receives messages $\perp \cdot x$ in σ'' when they would have received only x in profile (σ'_i, σ_{-i}) *and* x is interpreted as a valid message. When this occurs, the VER method of scheme π has failed. Thus, by the security of the secret sharing scheme π, the probability that this happens is at most ϵ_1 for each round in the game; it follows by the union bound that the total probability of the profiles differing is at most $n^2\epsilon_1$. Since utilities are in $[0,1]$, we have that for each player i, the expected utility of $(\sigma''_i, \sigma_{-i})$ is at most $n^2\epsilon_1$ less than (σ'_i, σ_{-i}). □

Deviation at a single round. In our next transformation, we restrict the first round in which an invalid message is sent by σ''_i. Let $\sigma'''_i(k)$ be the player i strategy that runs σ_i and σ''_i in parallel; if the two strategies differ in any round other than k (i.e., if σ'' starts appending $\perp$ to its messages), then $\sigma'''_i(k)$ continues following σ_i; if the difference occurs in round k, then it continues to run σ''_i. That is $\sigma'''_i(k)$ only starts to send invalid messages in round k. Among these n^2+2 profiles (recall that the game has n^2+1 rounds, and there is always the possibility of σ''_i never sending an invalid message), let m be the index of the profile with the greatest expected utility for player j, and let $\sigma'''_i = \sigma'''_i(m)$.

Claim 8. $u_i(\sigma'''_i, \sigma_{-i}) \geq u_i(\sigma) + \frac{\epsilon}{n^2+2} - \epsilon_1 > u_i(\sigma)$.

Proof: To prove the claim, let us consider the strategy τ_i that picks $k \in [n^2+2]$ at random and next runs $\sigma'''_i(k)$. By definition, $u_i(\sigma''', \sigma_{-i}) \geq u_i(\tau_i, \sigma_{-i})$. Furthermore, since τ_i picks k at random, we have that

$$u_i(\sigma''', \sigma_{-i}) \geq u_i(\tau, \sigma_{-i}) = \left(1 - \frac{1}{n^2+2}\right) u_i(\sigma) + \left(\frac{1}{n^2+2}\right) u_i(\sigma''_i, \sigma_{-i})$$

[6] Formally, this new message is not part of the message space of the game, so we actually need to consider a different game (with a potentially larger message space) where this message is valid. In fact, we might also have to enlarge the strategy space to make sure this strategy is valid.

This follows because τ selects a k for which σ_i'' deviates beginning at round k with probability $1/n^2 + 2$. Since this guess is independent of σ_i'', when it is correct, the expected utility is $u_i(\sigma_i'', \sigma_{-i})$. When the guess is incorrect, the utility is $u_i(\sigma)$. Finally, by Claim 7, we thus have that

$$u_i(\tau, \sigma_{-i}) \geq u_i(\sigma) + \frac{1}{n^2+2}\left(\epsilon - n^2\epsilon_1\right) \geq u_i(\sigma) + \frac{\epsilon}{n^2+2} - \epsilon_1 > u_i(\sigma)$$

□

We have thus reduced the deviation to a strategy σ_i''' which either explicitly cheats at round m or never deviates. We now consider the same two cases discussed in the RS proof.

Case 1: m is after the critical point. At least $t-2$ shares must have been broadcast in every row, and therefore other players who follow σ also broadcast their shares and guarantee that at least t legitimate shares in some row are broadcast. Thus, the remaining players run the reconstruct algorithm on the valid shares on this row and output the secret. This implies that the utility $u_i(\sigma''', \sigma_{-i})$ is at most $u_i(\sigma)$ when all players learn the secret, which contradicts Claim 8.

Case 2: m is before the critical point. The remaining players follow σ and therefore halt and do not output the secret. But, because m is prior to the critical point, at most $t-2$ shares in each row have been broadcast and thus player i may have at most $t-1$ shares for any row. By the security of the authenticated secret sharing scheme, it thus follows using exactly the same argument as in the RS proof that $u_j([\sigma, r, \sigma']_j) \leq u_j(\sigma)$, which contradicts Claim 8.

In both cases, we have shown that σ' cannot be an ϵ-deviation. □ □

5.3 Computational Rational Secret Sharing

As a toy example we show that the same protocol, but where the secret sharing scheme that is only computationally secure, is still an ϵ-RS if the strategy space consists of T-bounded strategies (instead of the full set of strategies). Roughly speaking, we say that a (t, n, ϵ)-authenticated secret-sharing scheme is (t, n, T, ϵ)-secure if it satisfies Definition 6, but 1) condition 1 is modified to only require that no T-bounded "distinguisher" can distinguish (subsets of shares) better than with probability ϵ (instead of requiring them to be identically distribution), and 2) condition 2 is only required to hold for T-bounded A (i.e., no T-bounded A can violate the authentication property with probability higher than ϵ). First note that it follows using a standard argument that the secret sharing presented is computationally secure if the underlying scheme is computationally secure (but at a polynomial degradation in the parameters; that is, there exists some polynomial p such that if the underlying scheme is $(t, n, p(T), p(\epsilon))$-secure, then the new one is only (t, n, T, ϵ)-secure.) We now turn to show that reconstruction strategy from the previous section is both ϵ-RS and ϵ-NE if the strategy space S consist of the set of T-bounded strategies. For this result, we require that our representation of circuit size is such that if a strategy σ is T-bounded, then the strategy $\hat{\sigma}$ which is identical to σ except that it does not send its last communication message is also T-bounded.[7]

We first note that Thm. 11 goes through unchanged for T-bounded strategies; we only need to check that all the strategies constructed in the proof still are T-bounded, which directly follows (by relying on the circuit representation assumption). Now, let us turn to the proof of Thm 14. Here, the only thing that needs to be verified is the reduction to the security of the secret-sharing scheme in Proposition 1 (in both case 1 and 2) and in Claim 6. In both cases, these reductions are polynomial-time and only degrade the

[7] In general, if σ is a T-bounded strategy, adding the logic so that σ skips the sending of the last message may result in σ no longer being T-bounded. However, this assumption can be easily satisfied, for instance, by requiring that every strategy is "wrapped" by an appropriate selection function that runs the original strategy and either prints a message or not depending on the round.

success probability polynomially, so there exists some polynomial p, such that if the secret-sharing scheme is $(t, n, p(T), p(\epsilon))$-secure, then the protocol is ϵ-RS and ϵ-NE when the strategy space is the set of T-bounded players.

Applications to secure function evaluation Finally, let us briefly consider the task of secure function evaluation (SFE) [**?**, **?**]. Here, we have a set of n player, each having their own private input x_i, that wish to compute a pre-specified function $f(\vec{x})$ of their inputs $\vec{x}$ by running a protocol π. Roughly speaking π is said to be secure if running it provides the same guarantees of "correctness" and "privacy" as if the players had been communicating directly with a trusted party (or mediator) that performs the computation for them. As shown in [HP10a], "perfect" cryptographic security guarantees that for any distribution over inputs for the players, if it is a NE for the players to provide their true inputs to the mediator (and output whatever the mediator tells them), then it is still a NE equilibrium for them to play the protocol using their inputs. If the protocol is only "computationally-secure", the protocol will only be an ϵ-NE for T-bounded players (where T and ϵ depend on the computational security of the protocol).

As we now argue, if we additionally consider 1) a notion of strictly competitive utilites, where instead of recovering the type of nature, the players try to retrieve the value $f(\vec{x})$ (where $\vec{x}$ are their types), and 2) a stronger notion of "collaboration is preferred" where all player i prefer everyone to get the output than to make a guess for the output even if the i knows any subset of $n-2$ inputs, then it seems that a simple combination of a traditional secure function evaluation protocol and our secret sharing protocol is ϵ-RS for bounded players, assuming 2 honest players. The protocol proceeds as follows: Run an SFE to compute a secret-sharing of the output; then run our reconstruction protocol. It is easy to see that the resulting protocol still satisfies the traditional cryptographic definition of security (if the original SFE protocol is secure). To argue that this protocol is RS, it seems that we can use the same approach as in Thm 14 and specifically Proposition 1:

- If a renegotiation takes place after the critical round, then as before, the value will be reconstructed with high probability.
- If a renegotiation takes place before the critical round, the two honest players stop talking. By the privacy property of the SFE protocol, it follows that at any such point, no coalition of $n-2$ players can guess the output with significantly better probability than before the protocol began.

We leave a formalization of this for future work. It is worthwhile to note that in order for the above protocol to be an RS, we only require that the SFE protocol in use is secure with respect to so-called "honest-but-curious" players, that honestly following the protocol instruction (but attempting to extract as much information as possible from the transcript). This a-priori seems surprising. But recall that our definition of RS only consider explicit renegotiations that are publicly announced by the renegotiation. On the other hand, to get a protocol that is also a NE, we require the SFE to be secure also with respect to so-called "malicious" players who does not necessarily follow the protocol instructions.

6 Future Work

Our work opens up for several interesting direction for future work.We outline some of these in this section.

Games with Unobserved actions A natural direction is to extend the notion of renegotiation-safety to handle games that have unobserved actions. The game in Fig. 7 is a classic example of a poor sequential equilibrium put forth by Kohlberg and Mertens [KM86]. The sequential equilibrium strategy is (L, a) with

beliefs $(1/3, 2/3)$ for player 2's information set. Intuitively, however, the (L, a) profile is not renegotiation-safe because Player 1 can renegotiate to (R, b). However, player 2 cannot observe whether player 1's first action was indeed R. In this case, there is no reason for player 1 to change, but in other cases, it might be beneficial for player 1 to announce a renegotiation to σ but actually make his unobserved move according to a different strategy. This aspect raises additional modeling questions.

An even more challenging question would be to try to model renegotiation in games with imperfect recall.

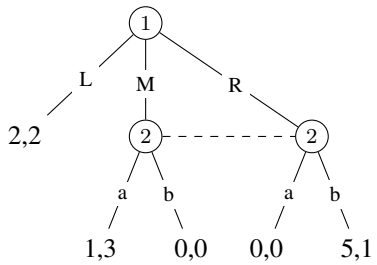

Figure 7: Kolhberg-Mertens example of a questionable sequential equilibrium

More on Renegotiations v.s. Deviations Even in games with perfectly observable actions, deviations might not always be easily detectable. For instance, when a player is supposed to be using randomized strategy, it might not be easy to detect that it deviated from it. More seriously, in the context of computationally bounded players, a player might not be able to detect that someone else sent an invalid message (that is not even in the support of the intended strategy; for instance, a comptuationally bounded player might not be able to check whether a large number is product of two or three primes). As argued in the introduction, in such models, RS alone provides weak guarantees against "hidden" deviations. Indeed, as argued, in such scenarios it is preferrable to consider both RS and some stability notion that considers secret deviations (for instance, such as NE as as we did in the context of secret-sharing). But even doing so does not necessarily protect against a renegotiation at a later round that started off by a hidden deviation (i.e., a player secretly starts to deviate in a way that is undetectable, and only later proposes a renegotiation). Dealing with such renegotiations seems interesting.

On modeling non-negotiating players In this paper we allowed the strategy of the honest player to depend on whether a renegotiation has taken place or not. It would be nicer if the strategy of honest players do not rely on this information; i.e., the strategy is simply a history from histories of actions to actions. (We conjecture that our secret-sharing protocol, in fact, still remains ϵ-RS for an appropriate choice of ϵ, even if the honest players decide whether to abort or not based on whether they have seen any invalid messages.)

More generally, it would be interesting to capture a notion of RS where renegotiations can done "secretly" among the participating players; that is, the renegotiating players are aware of the renegotiation, while the others are not (and furthermore, it is common knowledge among the renegotiators that the other players are not renegotiating). The coalition-safety guarantee in Ferreira's CPE notion [Fer95] seems relevant.

On the definition of ϵ-RS Our definition of ϵ-RS only requires that the renegotiation at the "first-level" should gain by more than ϵ in utility by renegotiating; further renegotiation need only gain by more than 0. As mentioned, the reason for this is that, if we had required that further renegotiations also gain by at least ϵ,

then RS would not necessarily imply ϵ-RS (as more renegotiations are now considered stable). On the other hand, if it is common knowledge that no one cares about an ϵ change in utility, this alternative model seems more appropriate. A middle ground would be to require that higher-level renegotiation require smaller and smaller change in utility (e.g., once a renegotiation has already been done, the cost of renegotiation goes down). Finding an appropriate model for dealing with this seems challenging.

Epistemic foundations of renegotiation Our formulation of RS begs for an epistemic characterization. Intuitively, it seems that our notion captures the intuition that it is common knowledge that players will renegotiate if they can gain from it. Thus, a player only renegotiates to a strategy if it is profitable, knowing that you might renegotiate it (and you know that the next person might renegotiate etc.).

References

[AAC+05] A Aiyer, L Alvisi, A Clement, M Dahlin, JP Martin, and C Porth. Bar fault tolerance for cooperative services. In *SOSP'05*, 2005.

[ADGH06] I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed Computing Meets Game Theory: Robust Mechanisms for Rational Secret Sharing and Multiparty Computation. In *PODC '06: Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 53–62, New York, NY, USA, 2006. ACM Press.

[Ash91] Geir B. Asheim. Extending renegotiation-proofness to infinite horizon games. *Games and Economic Behavior*, 3(3):278–294, August 1991.

[Ash97] Geir Asheim. Individual and collective time-consistency. *The Review of Economic Studies*, 64(3):427–443, July 1997.

[BK93] Jean-Pierre Benoît and Vijay Krishna. Renegotiation in finitely repeated games. *Econometrica*, 61(2):303–323, 1993.

[Blu87] A Blume. Renegotiation-proof theories in finite and infinite games. Technical report, University of California at San Diego, 1987.

[BM89] James Bergin and Bentley MacLeod. Efficiency and renegotiation in repeated games. Working Papers 752, Queen's University, Department of Economics, 1989.

[BP98] Elchanan Ben-Porath. Correlation without mediation: Expanding the set of equilibrium outcomes by "cheap" pre-play procedures. *Journal of Economic Theory*, 80(1):108–122, May 1998.

[BW87] B. Douglas Bernheim and Michael D. Whinston. Coalition-proof nash equilibria ii. applications. *Journal of Economic Theory*, 42(1):13–29, June 1987.

[Cav87] J Cave. Long term competition in a dynamic game: The cold fish dilemma. *Rand Journal of Economics*, 18:596–610, 1987.

[DBR89] B. Douglas Bernheim and Debraj Ray. Collective dynamic consistency in repeated games. *Games and Economic Behavior*, 1(4):295–326, December 1989.

[DeM88] P. DeMarzo. Coalitions and sustainable social norms in repeated games. Papers 529, Stanford - Institute for Thoretical Economics, 1988.

[Far83] Farrell. Credible repeated game equilibrium. 1983.

[Fer95] Jose Luis Ferreira. On the possibility of stable renegotiation. *Economics Letters*, 47(3-4):269–274, March 1995.

[FM87] J. Farrell and E. Maskin. Renegotiation in repeated games. Technical report, Harvard, 1987.

[GK06] S. Dov Gordon and Jonathan Katz. Rational secret sharing, revisited. In *SCN'06*, pages 229–241, 2006.

[GLR10] Ronen Gradwohl, Noam Livne, and Alon Rosen. Sequential rationality in cryptographic protocols. To appear in FOCS'10, 2010.

[HP10a] Joseph Y. Halpern and Rafael Pass. Algorithmic rationality: Game theory with costly computation. In *ICS'10*, 2010.

[HP10b] Joseph Y. Halpern and Rafael Pass. I don't want to think about it now: Decision theory with costly computation. In *KR'10*, 2010.

[HT04] Joseph Y. Halpern and Vanessa Teague. Rational secret sharing and multiparty computation: extended abstract. In *STOC'04*, pages 623–632, 2004.

[IML05] Sergei Izmalkov, Silvio Micali, and Matt Lepinski. Rational secure computation and ideal mechanism design. In *FOCS'05*, pages 585–595, 2005.

[KM86] Elon Kohlberg and Jean-Francois Mertens. On the strategic stability of equilibria. *Econometrica*, 54(5):1003–1037, 1986.

[KN08a] Gillat Kol and Moni Naor. Cryptography and game theory: Designing protocols for exchanging information. In *TCC'08*, pages 320–339, 2008.

[KN08b] Gillat Kol and Moni Naor. Games for exchanging information. In *STOC'08*, pages 423–432, 2008.

[KW82] D. Kreps and R. Wilson. Sequential equilibria. *Econometrica*, 50:863–894, 1982.

[LT06] Anna Lysyanskaya and Nikos Triandopoulos. Rationality and adversarial behavior in multiparty computation. In *CRYPTO'06*, pages 180–197, 2006.

[MS83] Roger Myerson and Mark Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(265–281), 1983.

[MS09] Silvio Micali and Abhi Shelat. Purely rational secret sharing. In *TCC'09*, 2009.

[OPRV09] Shien Jin Ong, David C. Parkes, Alon Rosen, and Salil P. Vadhan. Fairness with an honest minority and a rational majority. In *TCC*, pages 36–53, 2009.

[OR94] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT, 1994.

[vD89] Eric van Damme. Renegotiation-proof equilibria in repeated prisoner's dilemma. *Journal of Economic Theory*, 47(206–207), 1989.

A Prior work on Rational Secret Sharing

Halpern and Teague [HT04] consider the solution concept of iterated removal of weakly dominated strategies and first rule out any rational secret sharing protocol that terminates in a fixed number of rounds. They then suggest one (for $n = 3$ and later general n) that does not have a fixed upper bound on the number of rounds and relies on simultaneous-broadcast channels. A main limitation to the applicability of their protocol is that the dealer continues to be an active participant. (In most settings, such a dealer could directly inform the players of what the secret is.)

Gordon and Katz [GK06] present a protocol for $n = 2$ players which removes unwanted equilibria from the Halpern and Teague protocol, and dismisses the need for the periodic involvement of a trusted dealer. Their protocol too relies on the existence of *simultaneous broadcast* channels. Abraham, Dolev, Gonen, and Halpern [ADGH06] present a similar protocol, focussing on defining (and protecting against) coalitions of rational players.

Lysyanskaya and Triandopoulos [LT06] consider a model in which some players are rational, and some players are malicious. As in [GK06], their protocol uses an approach described by Ben-Porath [BP98] in the context of achieving correlated equilibrium. The idea is that each round of the protocol is either useful with probability β or a test-of-honesty with probability $1 - \beta$. At the beginning of the round, the players do not know which is the case, and thus have an incentive to behave honestly if β is chosen appropriately.

Kol and Naor [KN08a] present a different and insightful protocol enjoying a stronger and new notion of a Nash equilibrium: "everlasting equilibrium." In a follow-up work [KN08b], Kol and Naor present an information-theoretic protocol for which the honest strategy is a "strict Nash equilibrium", in essence a profile of strategies in which any player deviating from his own strategy expects to receive a strictly smaller utility if he believes that the other players will stick to their strategies. In all the above protocols the envisaged channels are simultaneous broadcast ones. The follow-up work also presents an ϵ-variant of the equilibrium when ordinary broadcast channels are used.

The works of Izmalkov, Lepinski and Micali [IML05] and Micali and shelat [MS09] achieve stronger notions of equilibria, but rely on physical assumptions (such as physical envelopes, and physical randomization devices) that provably cannot be implemented under standard communication channels.

Finally, Ong, Parkes, Rosen, and Vadhan [OPRV09] propose a protocol in a model quite different from prior work. Their protocol does not require any special channels: namely, they rely on ordinary broadcast channels (rather then simultaneous-broadcast ones). On the other, they restrict the rationality of a small number of players. Namely, they assume that $\omega(\log n)$ players are honest: that is, that they stick to their prescribed strategies no matter what. In this model, their protocol enjoys several nice properties. In particular, it yields a variant of the notion of perfect equilibrium.