K-NCC: Stability Against Group Deviations in Non-Cooperative Computation

Itai Ashlagi, Andrey Klinger, and Moshe Tennenholtz
Technion–Israel Institute of Technology
Haifa 32000, Israel

Abstract. A function is non-cooperative computable [NCC] if honest agents can compute it by reporting truthfully their private inputs, while unilateral deviations by the players are not beneficial: if a deviation from truth revelation can mislead other agents, then the deviator might end up with a wrong result. Previous work provided full characterization of the boolean functions which are non-cooperatively computable. Later work have extended that study in various directions. This paper extends the study of NCC functions to the context of group deviations. A function is K-NCC if deviations by a group of at most K agents is not beneficial: in order to mislead other agents, at least one group member might compute the wrong outcome. A function which is K-NCC for every K is termed *strong-NCC*. In this paper we provide a full characterization of the K-NCC functions, for every K, and of strong-NCC functions in particular. We show that the hierarchy of K-NCC functions is strict. Surprisingly, we also show that an anonymous function is NCC iff it is strong-NCC; that is, an anonymous function which is non-cooperatively computable is stable against deviations by any coalition of the agents. In addition, we show that group deviations are stable: if there exists a deviating coalition of minimal size K, then there is no sub-coalition of it which will benefit by further deviation from the original deviating strategy.

1 Introduction

Non-cooperative computing [NCC], introduced in [7], deals with the desire to compute a function defined on agents' private inputs where the agents might have incentives not to report truthfully. This can be viewed as a task of *informational mechanism design*. While in a classical mechanism design context (see [5] Chapter 23) the essence of the problem is the lack of information about the agents' preferences, in NCC the agents' preferences are known but other information they possess which is needed for the joint activity is private. NCC introduces a game-theoretic version of the problem of multi-party computation.[1].

In order to see the basic idea behind NCC consider for example the situation where each agent's secret is a bit, and the function to be computed is the parity function. If all agents report their bits honestly then the parity can be easily

[1] Indeed, the work in [4,1] deals with NCC when there is no center in the system, bridging the gap to the classical assumptions in the cryptographic and distributed computing literature

computed. However, if an agent reports 1 (resp. 0) instead of 0 (resp. 1), while all other agents report honestly, then this agent will be able to re-cover the true result by reversing the reported outcome, while misleading the other agents. Hence, the parity function is **not** non-cooperatively computable. On the other hand, if the function is the majority function, then false report might make the deviator unclear about the true result, given that the result of the majority function is computed and reported to the participants using a trusted center based on the information provided by them; this makes this function non-cooperatively computable.

The early results on NCC provided complete characterization of the functions which are non-cooperatively computable. Additional work has been carried out on extending this setting [6], as well as on considering the agents' costs, which lead to other forms of deviations [8].

In this paper we attack a major challenge: deviations by coalitions in the NCC setting. Although it has been already acknowledged that knowing whether a function is stable against deviations by groups of agents is central to the context of non-cooperative computing and rational multi-party computation [1], no analysis has been provided for the characterization of functions which are stable against deviations by coalitions in that context. While NCC is associated with honest computation being in equilibrium, group deviations in that context can be associated with the concept of strong equilibrium as introduced by Aumann [2]; therefore, we refer to a function as strong-NCC if no coalition can mislead in some cases at least one member which is not part of the coalition, without taking the risk this would cause at least one member of the coalition not to know the function value. More generally, we wish to study K-NCC functions, in which deviations of coalition of size at most K are considered. The case of NCC is then associated with 1-NCC functions.

In the NCC model there are n agents, each of which wish to compute an n-ary function w, with each of the agents holding one of the inputs to w. The process of computation is mediated by a center as follows: Each agent declares his input (truthfully or not) to the center, the center performs computation based on those inputs, and reports back to the agents an output. In the setting we deal with, the center applies w to the declared inputs and announces the value to all the agents. Each agent has now to decide on the output he accepts as a result of the computation.[2] We concentrate on agents whose utility function has two components. The main component, termed *correctness*, is the wish to compute the function correctly. The secondary component, termed *exclusivity*, is the wish that other agents do not compute the function correctly. The secondary

[2] In fact, under the famous revelation principle, one can show that the restriction to such mechanisms can be done without loss of generality.

component - exclusivity - is affecting the agent only if the main component - correctness - is not under risk. The definition of ”exclusivity” is rather loose, and allows for many possible variants; for example, it can be more important for an agent to mislead a particular agent than another agent. For the results presented in this paper the exact meaning of exclusivity does not matter; the only assumption is that the situation in which at least one other agent is mistaken is better than the situation where all other agents are correct, as long as the agent can compute the function correctly.

In the subsequent sections we present sound and complete conditions for a function to be K-NCC, and in particular strong NCC. We prove that an n-ary boolean function is Strong NCC (i.e. resilient to deviation of a coalition of any size) iff it is not dominated and not k-reversible for any $1 \leq k < n$. Our result implies that any anonymous function is strong NCC iff it is NCC. We also show that the hierarchy generated by K-NCC functions is strict and that when a function is not stable against deviation of a minimal coalition of size K, then such a deviation will be stable against further deviations of sub-coalitions. In game-theoretic terms, this result implies that the existence of a coalition-proof equilibrium implies the existence of a strong equilibrium in the NCC setting.

2 Definitions

In this section we define the notion of K-NCC. Given a set of agents $N = \{1, 2, \ldots, n\}$, and a special agent termed ”the center”, we assume that there exists a private secure communication line between every agent $i \in N$ and the center. The type v_i of agent i is selected from some domain B_i. We concentrate on a Boolean domain, where $B_i = B = \{0, 1\}$

Given a function $w : B^n \rightarrow B$, we consider the following protocol:

1. For any instantiated type vector $v \in B^n$, each agent i declares his type $\hat{v_i}$ to the center (truthfully or not; $\hat{v_i} = v_i$ may or may not hold).
2. The center computes the value $w(\hat{v}) = w(\hat{v_1}, \ldots, \hat{v_n})$ and announces it to all agents.
3. Each agent i computes $w(v)$ based on $w(\hat{v})$ and v_i (his true input).

The protocol defines a strategy space for each agent. A pure strategy for agent i is a pair of functions (f_i, g_i). $f_i : B \rightarrow B$, the *declaration function*, determines the input declared to the center based on the agent’s true input. The *truthful* declaration function is the identity function $f^t(v) = v$. $g_i : B^2 \rightarrow B$, the *interpretation function*, is used by the agent to decide on the value of the function based on the announcement by the center and his true input. The *trusting* interpretation function is the projection function $g^t(v_1, v_2) = v_1$ in which the agent simply accepts the value announced by the center. We will refer to the strategy (f^t, g^t) as the *straightforward* strategy.

Note that the strategy profile consisting only of straightforward strategies results in each agent computing w correctly for all input vectors. We are looking for functions for which such a strategy profile forms an equilibrium, and more generally a (k-)strong equilibrium which is stable against deviations of coalition (of size at most k). We will use the following notations:

Definition 1. *For a set of agents $C = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$, B_C is defined as $\prod_{j \in C} B_j$ and B_{-C} is defined as $\prod_{j \notin C} B_j$. In the same way, $v_C \in B_C$ is a tuple of types of agents participating in the set C, and $v_{-C} \in B_{-C}$ is a tuple of types of agents not participating in the set C.*

We can now define K-NCC:

Definition 2. *A function* w *is called K-NCC if the following holds: For any set of agents $C = \{i_1, \ldots, i_k\}$, $k \leq K$, every tuple of their strategies $((f_{i_1}, g_{i_1}), \ldots, (f_{i_k}, g_{i_k}))$, and every corresponding agent types $v_{i_j} \in B$, $1 \leq j \leq k$, it is the case that:*

- *either $\exists v_{-C} \in B_{-C}, \exists j, i_j \in C$, such that*

$$g_{i_j}(w(f_{i_1}(v_{i_1}), \ldots, f_{i_k}(v_{i_k}), v_{-C}), v_{i_j}) \neq w(v_{i_1}, \ldots, v_{i_k}, v_{-C})$$

- *or $\forall v_{-C} \in B_{-C}$ we have*

$$w(f_{i_1}(v_{i_1}), \ldots, f_{i_k}(v_{i_k}), v_{-C}) = w(v_{i_1}, \ldots, v_{i_k}, v_{-C})$$

In words, each deviating coalition of up to K players will either be mistaken for some types of the non-deviating players or will always produce the same result as if they didn't deviate

The following definition will play a key role in the characterization of functions which are K-NCC:

Definition 3. *A function* w *is called k-reversible if the following holds: $\exists C = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ such that $\forall v_{-C} \in B_{-C}, \forall j, 1 \leq j \leq k, \forall v_{i_j} \in B_{i_j}$,*

$$w(v_{i_1}, \ldots, v_{i_k}, v_{-C}) = 1 - w(1 - v_{i_1}, \ldots, 1 - v_{i_k}, v_{-C})$$

Note that in this definition, the set C contains exactly k players.

It may worth to notice that the definition of reversible functions discussed in [7] coincides with the definition of 1-reversible functions above.

Another definition, used in previous work on NCC is the following one:

Definition 4. *A function* w *is called dominated if the following holds:* $\exists i \in \{1,\ldots,n\}, v_i \in B$*, such that* $\forall v_{-\{i\}} \in B_{-\{i\}}, v'_{-\{i\}} \in B_{-\{i\}}$, $w(v_i, v_{-\{i\}}) = w(v_i, v'_{-\{i\}})$*, and there is some* $v_{-\{i\}} \in B_{-\{i\}}, v'_{-\{i\}} \in B_{-\{i\}}$*, for which* $w(1-v_i, v_{-\{i\}}) = 1 - w(1-v_i, v'_{-\{i\}})$.

If a function is dominated then, for a particular value of a particular agent's type, the agent knows the value of the function, while it can still influence the outcome by his report.

3 A Full Characterization of the K-NCC functions

Given the previous definitions, the characterization of NCC functions obtained in [7] can now be stated as follows:

Theorem 1. *A function is NCC iff it is not dominated and is not 1-reversible*

The following theorem establishes the exact conditions under which a function which is $(K-1)$-NCC is also K-NCC. It will imply necessary and sufficient conditions for a function to be K-NCC, and strong NCC.

Theorem 2. *A boolean function is* K*-NCC iff it is* $(K-1)$*-NCC and not k-reversible for* $k = K$.

Proof: $\Longrightarrow$ From the definition of K-NCC above, it follows that K-NCC$\subseteq (K-1)$-NCC. So a function that is not $(K-1)$-NCC can't be K-NCC.

Assume that the function *w* is k-reversible for $k = K$. Then there exist a set $C = \{i_1,\ldots,i_k\}$ such that for all $v_{-C} \in B_{-C}$, $w(v_{i_1},\ldots,v_{i_k},v_{-C}) = 1 - w(1-v_{i_1},\ldots,1-v_{i_k},v_{-C})$. Then each player i_j can use the following strategy:

$$(f_{i_j}(v) = 1 - v,\ g_{i_j}(r,v) = 1 - r)$$

meaning that he announces the opposite of his true type and interprets the opposite from what the center announced. Then, by *w* being k-reversible, all players $j, i_j \in C$ would be right, and all players which are not in C, playing truthfully, would be wrong. Thus a k-reversible function is not K-NCC.

$\Longleftarrow$ The opposite direction, the one that states that a function w that is (K-1)-NCC and not k-reversible for $k = K$ is K-NCC, will be proven by induction. The conditions under which a function is 1-NCC are described in[7], where it is shown that a function is 1-NCC iff it is not 1-reversible and not dominated. Notice that the requirement that a function is not dominated is not required for $K \geq 2$ since it is already implied by the fact the function is 1-NCC.

The base case: Assume that the function w is neither 1-reversible nor dominated. Consider agent i with strategy (f_i, g_i), and suppose all agents but i employ the straightforward strategy (f^t, g^t). Clearly, if i is irrelevant to w - that is, if $w(0, y_{-i}) = w(1, y_{-i})$ for all $y_{-i} \in B_{-i}$ - then $(f_i, g_i) = (f^t, g^t)$ is a best response for i. So assume that i is relevant, and assume further that $f_i \neq f^t$. Suppose agent i has the true input v_i, and declares value $f_i(v_i) = 1 - v_i$, and the center announces the value r. What could the value of $g_i(r, v_i)$ be? Since w is not dominated and since i is relevant, it cannot be that $g_i(r, v_i) = r$ without causing i to miscalculate for some input of the others. But at the same time it cannot be that $g_i(r, v_i) = 1 - r$, since this would imply that w is 1-reversible. From this contradiction it follows that necessarily $f_i = f^t$. But clearly if $f_i = f^t$ then $(f_i, g_i) = (f^t, g^t)$ is a best response for i (if all agent including i declare truthfully, i only loses by deviating from trusting interpretation function). That proves that if w is neither reversible nor dominated then it is 1-NCC.

Let $K \geq 2$, and assume that the function w is (K-1)-NCC and not k-reversible for $k = K$, but it is not K-NCC. Then there exists a deviating coalition $C = \{i_1, \ldots, i_k\}$ where $k = K$.

This coalition of deviating players C has a tuple of strategies $((f_{i_1}, g_{i_1}), \ldots, (f_{i_k}, g_{i_k}))$ which are better than the straightforward strategies. Therefore,

$$\exists v_{-C} \in B_{-C} \forall j \exists v_{i_j} \in B$$
$$w(v_{i_1}, \ldots, v_{i_k}, v_{-C}) \neq w(f_{i_1}(v_{i_1}), \ldots, f_{i_k}(v_{i_k}), v_{-C}) = r \tag{1}$$

That is, there is some combination of player types such that the deviating players succeed to mislead non-deviating players.

On the other hand:

$$\forall v_{-C} \in B_{-C}, \forall j, i_j \in C$$

$$g_i(w(f_{i_1}(v_{i_1}), \ldots, f_{i_K}(v_{i_K}), v_{-C}), v_{i_j}) = w(v_{i_1}, \ldots, v_{i_k}, v_{-C})$$

That is, none of the deviating players are ever mistaken (just as if they all played truthfully).

Consider one of the players in the deviating coalition $i_d \in C$.

If f_{i_d} is constant and w is not dependant on i_d's type, then any other deviating player $i_l \neq i_d$ can get correct result regardless of i_d's declaration. Hence, i_d could be ejected from the coalition and w wouldn't be (K-1)-NCC, contrary to the assumption, because a deviating coalition of size K-1 would exist.

If f_{i_d} is constant and w is dependant on i_d's type, then there exists $v_{-\{i\}} \in B_{-\{i\}}$ such that $w(0, v_{-\{i\}}) \neq w(1, v_{-\{i\}})$. Then other players in the deviating coalition C would have a possibility of mistake because they can't know whether to trust the center's announcement or not.

If i_d's declaration function is $f_{i_d}(v) = v$, i.e. the player is always saying the truth, then again he could be ejected from the coalition.

The only option we are left with for the declaration function of i_d is $f_{i_d}(v) = 1 - v$ for every $v \in B$. Thus, we get the following lemma:

Lemma 1. *The only possible declaration strategy for the players in a deviating coalition of size at least 2 is always to reverse their inputs.*

Let us now consider the possible interpretation functions of the player i_d. We are interested in $g_{i_d}(r, v_{i_d})$ for a specific value of v_{i_d}. According to (1), when the real type of the player i_d is v_{i_d} the center might output the value r, which is the wrong value of the function w. Then it must be that:

$$g_{i_d}(r, v_{i_d}) = 1 - r \text{ for the specific values } r, v_{i_d} \in B \tag{2}$$

On the other hand, if the real types of players in the deviating coalition are $1 - v_{i_j} = f_{i_j}(v_{i_j})$ then by lemma (1) the declared types are $f_{i_j}(f_{i_j}(v_{i_j})) = f_{i_j}(1 - v_{i_j}) = v_{i_j}$. According to (1), the center outputs the value $1 - r$ which is the wrong value of the function w. Then it must be that:

$$g_{i_d}(1 - r, 1 - v_{i_d}) = r \text{ for the specific values } 1 - r, 1 - v_{i_d} \in B \tag{3}$$

Now consider whether inequality (1) holds for *every* $v_C \in B_C, v_{-C} \in B_{-C}$. If it does, then this is precisely the condition for w to be k-reversible. However, this would be a contradiction to the initial assumption, so there exists $v'_C \in B_C, v'_{-C} \in B_{-C}$ such that:

$$w(v'_{i_1}, \ldots, v'_{i_k}, v'_{-C}) = w(f_{i_1}(v'_{i_1}), \ldots, f_{i_k}(v'_{i_k}), v'_{-C}) = r' \tag{4}$$

When the real type of the player i_d is v'_{i_d}, the center outputs r' which is the correct value of the function w. When the real type of the player i_d is $1 - v'_{i_d} = f_{i_d}(v'_{i_d})$, by lemma (1) the center outputs r' which is the correct value of the function w. Thus regardless of its type, player i_d must trust the center's announcement r':

$$g_{i_d}(r', v) = r' \text{ for the specific value of } r' \in B \text{ and any value of } v \in B \tag{5}$$

But equation 5 contradicts the set of equations (2 and 3). No interpretation function exists under such conditions.■

If we wouldn't assume that the function w is not k-reversible, then by lemma 1 all the deviating players would reverse their declared types and thus the output by the center would be the reverse of the true one, leading to the following lemma:

Lemma 2. *The only possible interpretation strategy for the players in a deviating coalition of size at least 2 is always to reverse the center's announcement.*

Theorem 2 does not deal with the case where $K = n$. As it turns out, this case is immediate:

Theorem 3. *A boolean function is n-NCC if it is (n-1)-NCC.*

Proof: Obviously a deviating coalition of n players cannot be beneficial to any of them, since there are no truthful player left to mislead.

We can now state the following necessary and sufficient conditions for a function to be K-NCC:

Corollary 1 *A boolean function w is K-NCC iff it is not dominated and not k-reversible for every $1 \leq k \leq K$. Therefore, a boolean function is strong NCC iff it is not dominated and not k-reversible for every $1 \leq k < n$.*

3.1 Anonymous Functions

An interesting class of functions are the *anonymous* functions (aka symmetric functions). The value of an anonymous Boolean function depends only on the number of 1's in the input. Many of the functions discussed in the computer science literature, such as parity, majority, consensus, order statistics, etc., are anonymous. As we now show for these functions any NCC function is also Strong NCC. This powerful result is implied by our characterization and the following theorem:

Theorem 4. *Anonymous n-variable functions that are not 1-reversible, are not k-reversible for any $1 < k < n$.*

Proof for even k: Assume w is an anonymous k-reversible boolean function and k is even. Then there exist a set $C = \{i_1, \ldots, i_k\}$, such that $w(x_i | i \in C, x_{-C}) = 1 - w(1 - x_i | i \in C, x_{-C})$ for every $x_{-C} \in B_{-C}$. Since k is even, lets take $k/2$ of those x_i to be 0, and $k/2$ of those to be 1. The function w is k-reversible, so like stated above $w(x_i | i \in C, x_{-C}) = 1 - w(1 - x_i | i \in C, x_{-C})$. On the other hand it is anonymous so if the number of 1's doesn't change, the result should not change. So $w(x_i | i \in C, x_{-C}) = w(1 - x_i | i \in C, x_{-C})$, which yields a contradiction.

Proof for odd k: Assume w is an anonymous boolean k-reversible function. $1 < k < n$ is odd and w is not 1-reversible. Then there exist a set $C = \{i_1, \ldots, i_k\}$ such that $w(x_i | i \in C, x_{-C}) = 1 - w(1 - x_i | i \in C, x_{-C})$ for every $x_{-C} \in B_{-C}$. The function w is anonymous, so its value depends only on the number of ones. Define $w'(t) = w(\text{Exactly t of the variables have value of } 1)$. Wlog assume that

$w'(0) = 0$.
$\Rightarrow w'(0) = 0 \quad \Rightarrow$ (Change 0 variables to '0', change k variables to '1')
$\Rightarrow w'(k) = 1 \quad \Rightarrow$ (Change 1 variable to '1', change $k-1$ variables to '0')
$\Rightarrow w'(2) = 0 \quad \Rightarrow$ (Change 2 variables to '0', change $k-2$ variables to '1')
$\Rightarrow w'(k-2) = 1 \Rightarrow$ (Change 3 variables to '1', change $k-3$ variables to '0')
...
It follows that $w = 0$ for even number of ones and $w = 1$ for odd number of ones. But that makes w 1-reversible contrary to the assumption.■

It now follows:

Corollary 2 *An anonymous n-variable Boolean function is Strong NCC iff it is NCC.*

This follows from Theorem 2 and Theorem 4 for coalitions smaller than n, and from Theorem 3 for coalitions of size n.

4 The K-NCC Hierarchy

The previous section established the characterization of K-NCC functions. We now show that the hierarchy implied by these functions is strict. To show this, we prove the following:

Theorem 5. *For $n > 2$, $K \leq n$, there exists a function w that is K-reversible, but not k-reversible for any $k < K$*

Proof: By construction:

For $K = 1$ it is enough to show that 1-reversible functions exist. One such function is the parity function.

For $K = 2$ consider the following function:

$$w(v_1, \ldots v_n) = \begin{cases} v_1 & \text{if } XOR(v_3, \ldots, v_n) = 0 \\ v_2 & \text{if } XOR(v_3, \ldots, v_n) = 1 \end{cases}$$

This function is 2-reversible by coalition $C = \{1, 2\}$. It is not 1-reversible by any single player.

For $K > 2$ consider the following function:

$$w(v_1, \ldots v_n) = \begin{cases} AND(v_1, \ldots, v_{K-1}) & \text{if } XOR(v_K, v_{K+1}, \ldots, v_n) = 0 \\ OR(v_1, \ldots, v_{K-1}) & \text{if } XOR(v_K, v_{K+1}, \ldots, v_n) = 1 \end{cases}$$

This function is K-reversible by coalition $C = \{1, \ldots, K\}$. It is left to show that it isn't k-reversible for some $k < K$. Assume w is k-reversible. Then let C be a smallest coalition for which w is $|C|$-reversible (we choose one of the smallest coalitions, if there are several coalitions of the same size.) Either there is exactly one player v_i, $i \geq K$ that participates in the reversing coalition C or there are no such players at all. If there were more reversing players with numbers at least K, two of them could be ejected from the coalition C, and the function wouldn't change. Contradiction to C being the smallest coalition.

Case 1. Such v_i participates: Consider $w(v_1 = 1, \ldots, v_{K-1} = 1, v_K = 0, \ldots, v_n = 0) = 1$. Coalition players reverse their inputs and v_i is one of them. $XOR(v_K, \ldots v_n)$ is now 1, so for w to be 0, $v_1 \ldots v_{K-1}$ must *all* change their input. This is a contradiction to the coalition being smaller than K.

Case 2. Such v_i doesn't participate: Consider the case where $v_K = 0, \ldots, v_n = 0$. Coalition players reverse their inputs and no v_i, $i \geq K$ is one of them. So w is effectively an AND function, and thus it is not reversible. ■

Theorems 2, 3 and 5 imply the strictness of the K-NCC hierarchy:

Corollary 3 *For $n \geq 2$ we have: n-NCC = (n-1)-NCC $\subset$ (n-2)-NCC $\subset \ldots \subset$ 2-NCC $\subset$ NCC*

5 Stability of Deviations

Stability against deviations by coalitions is typically considered as a very demanding requirement. Nevertheless, we have shown that for anonymous functions a function is stable against deviations by coalitions iff it is stable against unilateral deviations. This can be viewed as a highly positive result. When considering more general situations, it is natural to consider the question of whether the deviations themselves are stable. Indeed, this has led to the introduction of solution concepts, such as coalition-proof equilibrium [3], in which the stability of the deviations is considered. In this section we show that our results remain the same when considering only stable deviations. This will be proven in a very general setup. Namely, we will show that if there exist a deviating coalition, then there exist a stable deviating coalition. Formally, we will consider minimal deviations, and show that any further deviation by a smaller sub-coalition is not beneficial. This implies that in the NCC setting the existence of coalition proof equilibrium [3] coincides with the existence of strong equilibrium [2] (which by itself coincides with the existence of NCC for anonymous functions).

Theorem 6. *Consider a minimal deviation from the straightforward strategy profile, by a coalition C. Then, no further deviation by $C' \subset C$ can be beneficial to C'.*

Proof: From lemmas 1 and 2 it follows that the only possible deviation from straightforward strategies, for a coalition of size 2 or more, is if all players in the deviating coalition reverse their declared input and output the reverse of the center's announcement. Notice that in this case, all non-deviating players are always wrong. Let C be the smallest worthwhile deviating coalition. For C to be not self-enforcing, a new coalition $C' \subset C$ must be worth forming. So suppose a new coalition $C' \subset C$ is worthwhile.

In this situation, let us call it "Scenario1", the declaration and interpretation functions are these:

$$\begin{array}{lll}
f_{C \wedge \overline{C'}}(v) = 1 - v & g_{C \wedge \overline{C'}}(r, v) = 1 - r & \text{By original deviation} \\
f_{C \wedge C'}(v) = f'_{C \wedge C'}(v) & g_{C \wedge C'}(r, v) = g'_{C \wedge C'}(r, v) & f', g'\text{are some functions} \\
f_{\overline{C}}(v) = f^t & g_{\overline{C}}(r, v) = g^t & \text{Playing truthfully}
\end{array}$$

Players in C' are *never* mistaken and players in $\overline{C'}$ are *sometimes* mistaken.

Now consider "Scenario2", the following tuple of strategies:

$$\begin{array}{ll}
f_{C \wedge \overline{C'}}(v) = v = f^t & g_{C \wedge \overline{C'}}(r, v) = r = g^t \\
f_{C \wedge C'}(v) = 1 - f'_{C \wedge C'}(v) & g_{C \wedge C'}(r, v) = 1 - g'_{C \wedge C'}(r, v) \\
f_{\overline{C}}(v) = f^t & g_{\overline{C}}(r, v) = g^t
\end{array}$$

Every player not in the new coalition is playing truthfully, and the new coalition is playing in some unknown way, but every player in C is declaring the opposite from what he would declare in Scenario1.

Note, that between Scenarios 1 and 2 *all* the players in C (and just them) reversed their declared type. Since w is $|C|$-reversible this means that the center's announcement also reversed: $r_{scenario1} = 1 - r_{scenario2}$. Since also the interpretation function $g_{C'}$ of all players in C' in Scenario2 is opposite to Scenario1, and in Scenario1 all players in C' were correct, that means that in Scenario2 all players in C' are correct.

There could be two cases why the second deviation by C' is worthwhile. Either to mislead a player in $C \wedge \overline{C'}$, who was *always correct* after the first deviation by C, or to help a player in $\overline{C}$ who was *always wrong*, or both of them together.

In the first case, in Scenario1 some player in $C \wedge \overline{C'}$ had a possibility of mistake. Then in Scenario2, since both the interpretation function $g_{C \wedge \overline{C'}}$ and the center's announcement were reversed, some player in $C \wedge \overline{C'}$ still has a possibility of mistake.

In the second case, in Scenario1 some player in $\overline{C}$ has a possibility of being correct. Then in Scenario2, since the interpretation function $g_{\overline{C}}$ stayed the same - g^t, and the center's announcement was reversed, that player in $\overline{C}$ has a possibility of mistake.

In either of these cases, all players in C' are always correct and there is a player in $\overline{C'}$ that has a possibility of mistake. Thus C' can make a worthwhile deviation from the straightforward strategy $(1 - f', 1 - g')$. Remember, that $|C'| < |C|$ in contradiction to C being the smallest worthwhile coalition.■

6 Some words on computability

Although a simple rule is given to determine whether a function is K-NCC for some K, it is computationally hard to check those conditions.

Theorem 7. *Determining whether a function is* dominated *is NP-hard.*

Proof by reduction from SAT: Assume there is a polynomial algorithm that receives a function f and determines whether it is dominated. Then the SAT of a function $s(v_1, \ldots, v_n)$ can be solved in the following way:

Construct $f(v_1, \ldots, v_n, v_{n+1}) = \begin{cases} 1 & \text{if } v_{n+1} = 1 \\ s(v_1, \ldots, v_n) & \text{if } v_{n+1} = 0 \end{cases}$

If f is dominated, this would mean that f is dominated by variable v_{n+1}. This is so, because if f is a tautology it is not dominated by definition of dominated functions, and if f is not a tautology, it's outcome is dependant on v_{n+1}. f is dominated by v_{n+1} iff when $v_{n+1} = 0$, s is not satisfiable. Which means that f is dominated iff s is not satisfiable. ∎

Theorem 8. *Determining whether a function is* 1-reversible *is NP-hard.*

Proof by reduction from SAT: Assume there is a polynomial algorithm that receives a function f and determines whether it is 1-reversible. We can even assume that the algorithm only works for non-dominated functions. Then the SAT of a function $s(v_1, \ldots, v_n)$ can be solved in the following way:

Construct $f(v_1, \ldots, v_n, v_{n+1}, v_{n+2}) = \begin{cases} 1 & \text{if } (v_{n+1} \oplus v_{n+2}) = 1 \\ s(v_1, \ldots, v_n) & \text{if } (v_{n+1} \oplus v_{n+2}) = 0 \end{cases}$

If f is 1-reversible, this would mean that f is 1-reversible by either of the variables v_{n+1} and v_{n+2}. This is so, because reversing any other variable without changing v_{n+1} or v_{n+2} won't necessarily change f. f is 1-reversible iff s is not satisfiable. ∎

Note, that v_{n+2} was added in the above proof only to make f non-dominated. Also in an analogue way the same result can be proven for K-reversibility for any K.

References

1. Ittai Abraham, Danny Dolev, Rica Gonen, and Joe 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.
2. R.J. Aumann. Acceptable points in general cooperative n-person games. In A.W. Tucker and R.D. Luce, editors, *Contribution to the Theory of Games, Vol. IV, Annals of Mathematics Studies, 40*, pages 287–324. 1959.

3. B. Douglas Bernheim, Bezalel Peleg, and Michael D. Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, June 1987. available at http://ideas.repec.org/a/eee/jetheo/v42y1987i1p1-12.html.
4. J. Halpern and V. Teague. Rational secret sharing and multiparty computation. In *Proc. of STOC 2004*, 2004.
5. A. Mas-Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*. Oxford University Press, 1995.
6. R. McGrew, R. Porter, and Y. Shoham. Towards a general theory of non-cooperative computation. In *Proc. of the 9th conference of theoretical aspects of rationality and knowlede (TARK 2003)*, pages 59–71, 2003.
7. Y. Shoham and M. Tennenholtz. Non-cooeartive computation: Boolean functions with correctness and exclusivity. *Theoretical Computer Science*, 343:97–113, 2005.
8. Rann Smorodinsky and Moshe Tennenholtz. Overcoming free riding in multi-party computations–the anonymous case. *Games and Economic Behavior*, 55(2):385–406, May 2006. available at http://ideas.repec.org/a/eee/gamebe/v55y2006i2p385-406.html.