A Formal Proofs

Here, we present the proofs of our formal results.

A.1 Proofs of Section 4 (Benaloh According to Nash)

Lemma 1. *If $s_V = [p_1^V, \cdots, p_{n_{max}}^V]$ and $s_D = [p_1^D, \cdots, p_{n_{max}}^D]$ form a Nash equilibrium, then for all $i = V, D$ and $n = 1, \ldots, n_{max}$ we have $p_n^i > 0$.*

Proof. Suppose that (s_V, s_D) is a Nash equilibrium, and that $p_n^V = 0$ for some n (i.e., the voter always audits in round n). Take the smallest such n. Then, $s_D = n$ is the unique best response of D, i.e., the device must cheat for the first time in that round. We consider two cases now: (i) $n = 1$: in that case, the voter is better off playing $s_V = 1$, i.e., casting deterministically at the first round. (ii) $n > 1$: in that case, the voter is better off by swapping p_{n-1}^V and p_n^V, i.e., postponing the action planned for round $n-1$ until round n. In both cases, we get that (s_V, s_D) is not a Nash equilibrium, which is a contradiction. Hence, we get that $p_n^V > 0$ for all n. [*]

Suppose now that $p_n^D = 0$ for some n (i.e., the device never cheats in round n). Take the smallest such n. If $n = 1$, then V's best response is $s_V = 1$, which contradicts [*]. If $n > 1$, then V's best response includes $p_{n-1}^V = 0$, i.e., V postpones casting at $n-1$ until the next round, which also contradicts [*]. Hence, also $p_n^D > 0$ for all n. □

Lemma 2. *If $s_V = [p_1^V, \cdots, p_{n_{max}}^V]$ is a part of Nash equilibrium then $p_{n+1}^V = \frac{Succ_D}{Succ_D + Fail_D}\, p_n^V$ for every $n \in \{1, \ldots, n_{max} - 1\}$.*

Proof. Recall Condition (2), saying that:

$$\forall n_{cheat}, n'_{cheat} \in \{1, \ldots, n_{max}\}\ .\ u_D(s_V, n_{cheat}) = u_D(s_V, n'_{cheat}).$$

It is equivalent to:

$$\forall n \in \{1, \ldots, n_{max} - 1\}\ \ .\ \ u_D(s_V, n+1) - u_D(s_V, n) = 0 \qquad [*]$$

Notice that:

$$\begin{aligned} u_D(s_V, n) &= \sum_{i=1}^{n_{max}} p_i^V \cdot u_D(i, n) = \\ &= \sum_{i=1}^{n-1} p_i^V \cdot 0 + p_n^V \cdot Succ_D + \sum_{i=n+1}^{n_{max}} p_i^V \cdot (-Fail_V) \\ &= Succ_D \cdot p_n^V - Fail_D \cdot \sum_{i=n+1}^{n_{max}} p_i^V \end{aligned}$$

Similarly,

$$u_D(s_V, n+1) = \sum_{i=1}^{n_{max}} p_i^V \cdot u_V(i, n+1) =$$
$$= Succ_D \cdot p_{n+1}^V - Fail_D \cdot \sum_{i=n+2}^{n_{max}} p_i^V$$

By this and [*], we get that:

$$Succ_D \cdot p_{n+1}^V - Succ_D \cdot p_n^V + Fail_D \cdot p_{n+1}^V \;=\; 0$$

In consequence,

$$p_{n+1}^V = \frac{Succ_D}{Succ_D + Fail_D} p_n^V$$

which completes the proof. □

Theorem 1 *The mixed voting strategy* $s_V = [p_1^V, \cdots, p_{n_{max}}^V]$ *is a part of Nash equilibrium iff, for every* $n \in \{1, \dots, n_{max}\}$*:*

$$p_n^V \;=\; \frac{(1-R)R^{n-1}}{1-R^{n_{max}}}, \qquad \textit{where } R = \frac{Succ_D}{Succ_D + Fail_D}.$$

Proof. If s_V is a part of Nash equilibrium then $p_n^V > 0$ for all $n = 1, \dots, n_{max}$ (by Lemma 1). Moreover, by Lemma 2, the probabilities $p_1^V, \dots, p_{n_{max}}^V$ form a geometric sequence with ratio $R = \frac{Succ_D}{Succ_D + Fail_D}$. Thus, $\sum_{n=1}^{n_{max}} p_n^V = p_1^V \cdot \frac{1-R^{n_{max}}}{1-R}$ must be equal to 1. In consequence, $p_1^V = \frac{1-R}{1-R^{n_{max}}}$, and hence $p_n^V = \frac{(1-R)R^{n-1}}{1-R^{n_{max}}}$.

Notice that the above probability distribution is the only admissible solution, i.e., no other s_V can be a part of Nash equilibrium. By Nash's theorem, the finite Benaloh game must have at least one equilibrium; hence, it is the unique one. □

Theorem 2 *The behavioral voting strategy* $b_V = [b_1^V, \cdots, b_{n_{max}}^V]$ *is a part of Nash equilibrium iff, for every* $n \in \{1, \dots, n_{max}\}$*:*

$$b_n^V \;=\; \frac{1-R}{1-R^{n_{max}-n+1}}, \qquad \textit{where } R = \frac{Succ_D}{Succ_D + Fail_D}.$$

Proof. We claim that the above behavioral strategy implements the unique Nash equilibrium strategy $s_V = [p_1^V, \dots, p_{n_{max}}^V]$ of Theorem 1. To prove this, it suffices to verify that $p_n^V = (1-b_1^V) \cdot \dots \cdot (1-b_{n-1}^V) \cdot b_n^V$ for all $n = 1, \dots, n_{max}$. That is, casting at round n indeed corresponds to unsuccessful Bernoulli trials in the first $n-1$ rounds, and a successful trial in round n. The check is technical but straightforward. □

Theorem 3 *If* $\frac{Succ_D}{Fail_D} \to 0$*, then the equilibrium strategy* b_V *of the voter converges to the following behavioral strategy:*

$$\widehat{b_n^V} \;=\; \begin{cases} \frac{Fail_D}{Succ_D+Fail_D} & \text{for } n < n_{max} \\ 1 & \text{for } n = n_{max} \end{cases}$$

Proof. Take the behavioral NE strategy b_V in Theorem 2. For $\frac{Succ_D}{Fail_D} \to 0$, we get $R \to 0$. Hence, $1 - R^{n_{max}-n+1}$ for $n < n_{max}$ converges to 1 much faster than $1-R$, and thus $b_n^V = \frac{1-R}{1-R^{n_{max}-n+1}}$ gets arbitrarily close to $1 - R = \frac{Fail_D}{Succ_D+Fail_D}$. □

Theorem 4 *For* $n_{max} = 2$*, the behavioral NE strategy of the voter is:*

$$b_1^V \;=\; \frac{Succ_D + Fail_D}{2Succ_D + Fail_D}, \qquad\qquad b_2^V \;=\; 1.$$

Proof. Fix $n_{max} = 2$. By Theorem 2, we get $b_1^V = \frac{1-R}{1-R^2} = \frac{1}{1+R} = \frac{Succ_D+Fail_D}{2Succ_D+Fail_D}$. Similarly, $b_2^V = \frac{1-R}{1-R} = 1$. □

A.2 Proofs of Section 5 (Benaloh According to Stackelberg)

Lemma 3. *The best response of the device to any fixed strategy of the voter is*

$$BR_D(p^V) = \begin{cases} 0 & \text{for } p^V < p^V_{NE} \\ 1 & \text{for } p^V > p^V_{NE} \\ \text{any } p^D \in [0,1] & \text{for } p^V = p^V_{NE} \end{cases}$$

where $p^V_{NE} = \frac{Succ_D+Fail_D}{2Succ_D+Fail_D}$ *is the NE probability of casting in round 1.*

Proof. Given a strategy profile represented by (p^V, p^D), the expected payoff of the device is:

$$\begin{aligned} Eu_D(p^V, p^D) &= p^V p^D Succ_D - (1-p^V)p^D Fail_D + (1-p^V)(1-p^D)Succ_D \\ &= (2p^V Succ_D + p^V Fail_D - Succ_D - Fail_D)p^D + (1-p^V)Succ_D. \end{aligned}$$

Therefore, the derivative of $Eu_D(p^V, p^D)$ is

$$\frac{dEu_D(p^V, p^D)}{dp^D} \;=\; 2p^V Succ_D + p^V Fail_D - Succ_D - Fail_D,$$

which is negative for $p^V < \frac{Succ_D+Fail_D}{2Succ_D+Fail_D}$ and positive for $p^V > \frac{Succ_D+Fail_D}{2Succ_D+Fail_D}$. We recall from Theorem 4 that $p^V_{NE} = \frac{Succ_D+Fail_D}{2Succ_D+Fail_D}$ is the Nash equilibrium probability that the voter casts in the first round.[4] Thus, $Eu_D(p^V, p^D)$ is decreasing for $p^D \in [0, p^V_{NE})$, and hence reaches its maximum at $p^D = 0$. Similarly, $Eu_D(p^V, p^D)$ is increasing for $p^D \in (p^V_{NE}, 1]$, and has its maximum at $p^D = 1$.

Finally, by Lemma 1 and the necessary Nash condition (2), any response of D to strategy represented by p^V_{NE} must obtain the same expected payoff for D, hence each is a best response. □

Lemma 4. *The voter's expected utility against best response is:*

$$Eu_V(p^V, BR_D(p^V)) = \begin{cases} p^V Succ_V - (1-p^V)(c_{audit} + Fail_V) & \text{for } p^V < p^V_{NE} \\ -p^V Fail_V - (1-p^V)c_{audit} & \text{for } p^V \geq p^V_{NE} \end{cases}$$

Proof. For $p^V < p^V_{NE}$, we have $Eu_V(p^V, BR_D(p^V)) = Eu_v(p^V, 0) = p^V Succ_V - (1-p^V)(c_{audit} + Fail_V)$. Similarly, for $p^V > p^V_{NE}$, we have $Eu_V(p^V, BR_D(p^V)) = Eu_v(p^V, 1) = -p^V Fail_V - (1-p^V)c_{audit}$.

For $p^V = p^V_{NE}$, any $p^D \in [0,1]$ is a best response. Since $Eu_V(p^V_{NE}, p^D)$ is a linear function w.r.t. p^D, it reaches its minimum for either $p^D = 0$ or $p^D = 1$. Observe that $Eu_V(p^V_{NE}, 0) - Eu_V(p^V_{NE}, 1) = (2Fail_V + Succ_V)p^V_{NE} - Fail_V > 0$ because $p^V_{NE} = \frac{Succ_D + Fail_D}{2Succ_D + Fail_D} > \frac{1}{2} > \frac{Fail_V}{2Fail_V + Succ_V}$. Thus, $Eu_V(p^V_{NE}, 0) > Eu_V(p^V_{NE}, 1)$, and V's lowest payoff against best response at p^V_{NE} is $Eu_V(p^V_{NE}, 1)$. □

Theorem 5 *The following properties hold for the Benaloh game with $n_{max} = 2$:*

1. *There is no Stackelberg equilibrium for V in randomized strategies.*
2. *The Stackelberg value of the game is $SVal_V = \frac{Succ_D(Succ_V - Fail_V - c_{audit}) + Fail_D Succ_V}{2Succ_D + Fail_D}$.*
3. *$SVal_V > Eu_V(p^V_{NE}, p^D_{NE})$, where (p^V_{NE}, p^D_{NE}) is the Nash equilibrium.*
4. *If $Fail_D \gg Succ_D$ and $Succ_V \geq aFail_V$ for a fixed $a > 0$, then $SVal_V > 0$.*

Proof. Ad. 1 & 2: Consider $f(p^v) = Eu_V(p^V, BR_D(p^V))$, established in Lemma 4. The function is increasing for $p^V \in [0, p^V_{NE})$ and decreasing for $p^V \in [p^V_{NE}, 1]$. Moreover, $\lim_{p^V \to (p^V_{NE})^-} f(p^V) = Eu_V(p^V_{NE}, 0) > Eu_V(p^V_{NE}, 1) = f(p^V_{NE}))$. Thus, $SVal_V = \sup_{p^V \in [0,1]} f(p^V) = Eu_V(p^V_{NE}, 0) = p^V_{NE} Succ_V - (1 - p^V_{NE})(c_{audit} + Fail_V) = \frac{(Succ_D + Fail_D)(Succ_V + Fail_V + c_{audit})}{2Succ_D + Fail_D}$, and the value is not reached by any p^V. □

Ad. 3: By Lemma 1, $p^D_{NE} > 0$. Moreover, $Eu_V(p^V_{NE}, p^D)$ is linear w.r.t. p^D, and we already know that $Eu_V(p^V_{NE}, 0) > Eu_V(p^V_{NE}, 1)$, thus it must be strictly decreasing. In consequence, $SVal_V = Eu_V(p^V_{NE}, 0) > Eu_V(p^V_{NE}, p^D_{NE})$. □

Ad. 4: Let $Succ_V \geq aFail_V$, and recall that $c_{audit} < Fail_V$. Then, $SVal_V \geq \frac{Succ_D(aFail_V - Fail_V - Fail_V) + aFail_D Fail_V}{2Succ_D + Fail_D} = Fail_V(a - \frac{(2+a)Succ_D}{2Succ_D + Fail_D})$. For $\frac{Succ_D}{Fail_D} \to 0$, this converges to $aFail_V$, which is greater than 0. □

[4] Note that, for $n_{max} = 2$, mixed and behavioral strategies coincide and can be used interchangeably.