Strategic Disclosure of Collective Actions: Supplementary Appendix
(Not for Publication)

Geoffroy de Clippel* Kfir Eliaz†

This Version: July 2012

1 Proof of Equation (6) in the Main Paper

Notice that $\sigma_{-i}(\epsilon) \geq \sigma_{-i}(\eta)$ for $\epsilon \leq \eta$, and hence $ENG_i(x, \sigma^{\epsilon}_{-i}) \leq ENG_i(x, \sigma^{\eta}_{-i})$ by the third condition of Lemma 1. In turn, the right-hand side is lower or equal to $ENG_i(\eta, \sigma^{\eta}_{-i})$ for $x \leq \eta$, by the first condition of Lemma 1. Hence we will be done proving (6) after showing that $ENG_i(\eta, \sigma^{\eta}_{-i}) < 0$. We prove this inequality for $i = 1$. The case $i = 2$ follows from a symmetric argument. The first individual's expected net gain of disclosing when of type η is $\eta F(\eta) + \int_{y=\eta}^{1}(r_1(\eta, y) - (1-y))f(y)dy$. Suppose η is very small. The integral can be split into an integral for $y \in [\eta, 1-\eta]$, in which case the integrand is non-positive, and an integral for $y \in [1-\eta, 1]$, in which case the integrand is non-negative. The former term (for $y \in [\eta, 1-\eta]$) is smaller or equal to $\int_{y=\eta}^{1/2}(r_1(\eta, y) - (1-y))f(y)dy$, which itself is smaller or equal to $\int_{y=\eta}^{1/2}(1/2 - (1-y))f(y)dy$ since $r_1(\eta, y) \leq 1/2$, for all $y \in [\eta, 1/2]$ (by regularity). The latter term (for $y \in [1-\eta, 1]$) is smaller or equal to $\int_{y=1-\eta}^{1}(\eta - (1-y))f(y)dy$. To summarize, the first individual's expected net gain of disclosing when of a small type η is smaller or equal to $\int_{y=\eta}^{1/2}(1/2 - (1-y))f(y)dy + \int_{y=1-\eta}^{1}(\eta - (1-y))f(y)dy$. Notice that this expression is

*Department of Economics, Brown University. Email: declippel@brown.edu. Financial support from the Deutsche Bank through the Institute for Advanced Study is gratefully acknowledged.

†Department of Economics, Brown University. Email: kfir_eliaz@brown.edu

continuous in x, and strictly negative at $\eta = 0$. Hence there must exist $\eta > 0$ small enough for which it remains negative, and we are done proving (6). ■

2 Examples

In this section, we compute the equilibrium outcomes when types are uniformly distributed on the line joining $(1, 0)$ to $(0, 1)$, in the case of the coin-flip rule, and the bargaining solutions of Nash and Kalai-Smorodinsky.

2.1 Coin-Flip

Recall from the proof of Proposition 1 that the best response to any strategy is a threshold strategy, and hence one may restrict attention to best responses in terms of the thresholds. Because the coin-flip rule is continuous, the first individual's best response threshold θ_1 as a function of the second individual's threshold θ_2 is obtained by looking for the root of the first individual's expected net gain function (see footnote 7 in the main paper):

$$ENG_1^{r^{CF}}(\theta_1, \sigma_2^{\theta_2}) = \theta_1\theta_2 + \int_{y=\theta_2}^{1} \frac{\theta_1 - (1-y)}{2} dy = 0$$

or

$$\frac{\theta_1 + \theta_2 + \theta_1\theta_2}{2} - \frac{1+\theta_2^2}{4} = 0$$

which gives for $i = 1, 2$ and $j \neq i$:

$$\theta_i = BR_i(\theta_j) = \frac{(1-\theta_j)^2}{2(1+\theta_j)}$$

The static disclosure game thus admits a unique BNE, which is the symmetric equilibrium with common threshold $-2 + \sqrt{5} \sim 0.236$. Proposition 2 applies, and the unique BNE is thus the only profile of strategies that are rationalizable.

As for the dynamic disclosure game, the threshold is given by (see Proposition 7):

$$\int_{y=\theta_D^{CF}}^{1} \frac{\theta_D^R - (1-y)}{2} dy = 0$$

which yields $\theta_D^{CF} = \frac{1}{3}$.

2.2 Kalai-Smorodinsky

The threshold for the static disclosure game is given by:

$$\theta_{KS}^2+\int_{y=\theta_{KS}}^{1-\theta_{KS}}[\frac{1-y}{1-\theta_{KS}+1-y}-(1-y)]dy+\int_{y=1-\theta_{KS}}^{1}[\frac{\theta_{KS}}{\theta_{KS}+y}-(1-y)]dy=0.$$

Re-arranging, developing, and making the change of variables $z=1-y$ in the first part of the second term yields:

$$\theta_{KS}^2-\int_{y=\theta_{KS}}^{1}(1-y)dy+\int_{z=\theta_{KS}}^{1-\theta_{KS}}\frac{z}{1-\theta_{KS}+z}dz+\int_{y=1-\theta_{KS}}^{1}\frac{\theta_{KS}}{\theta_{KS}+y}dy=0.$$

Using integration by parts, this equation reduces to[1]

$$\frac{3\theta_{KS}^2-2\theta_{KS}+1}{2}-(1-\theta_{KS})ln(2-2\theta_{KS})-\theta_{KS}^2 ln(1+\theta_{KS})=0.$$

Solving this equation numerically yields that θ_{KS} is approximately 0.22 (for the symmetric BNE).

As for the dynamic disclosure game, the threshold is given by:

$$\int_{y=\theta_D^{KS}}^{1-\theta_D^{KS}}[\frac{1-y}{1-\theta_D^{KS}+1-y}-(1-y)]dy+\int_{y=1-\theta_D^{KS}}^{1}[\frac{\theta_D^{KS}}{\theta_D^{KS}+y}-(1-y)]dy=0$$

which yields $\theta_D^{KS}\approx 0.34$.

2.3 Nash

The threshold in the static disclosure game is given by:

$$ENG_1^{r_N}(\theta_1,\sigma_2^{\theta_2})=\theta_1\theta_2+\int_{y=\theta_2}^{1/2}[\frac{1}{2}-(1-y)]dy+\int_{y=1/2}^{1-\theta_1}0dy+\int_{y=1-\theta_1}^{1}[\theta_1-(1-y)]dy=0$$

or

$$\frac{\theta_1^2}{2}+\theta_1\theta_2+\frac{\theta_2}{2}-\frac{\theta_2^2}{2}-\frac{1}{8}=0$$

which gives for $i=1,2$ and $j\neq i$:

$$\theta_i=BR_i(\theta_j)=-\theta_j+\sqrt{2\theta_j^2-\theta_j+\frac{1}{4}}$$

[1]Integrating by parts, one gets $\int\frac{w}{\alpha+w}=w-\alpha ln(\alpha+w)$, for each α such that $\alpha+w>0$. Hence the sum of the third and fourth terms is equal to $[z-(1-\theta)ln(1-\theta+z)]_{z=\theta}^{1-\theta}+\theta[y-\theta ln(\theta+y)]_{y=1-\theta}^{1}$, or $1-2\theta-(1-\theta)ln(2-2\theta)+\theta[\theta-\theta ln(1+\theta)]$.

One can thus conclude that the disclosure game admits three BNEs, two in which one individual systematically discloses his type while the other discloses his type only when it falls above 1/2, and the unique symmetric equilibrium where the common threshold equals $(-1+\sqrt{3})/4 \sim 0.183$.

As for the dynamic disclosure game, the threshold is given by:

$$\int_{y=\theta_D^N}^{1/2} [\frac{1}{2} - (1-y)]dy + \int_{y=1-\theta_D^N}^{1} [\theta_D^N - (1-y)]dy = 0$$

which yields $\theta_D^N = \frac{1}{4}$.

3 Proof of Proposition 2

Let Σ be the set of strategies, for either individual,[2] that survive the iterated elimination of strictly dominated strategies. Let then $\theta = \sup\{x \in [0,1] | (\forall \sigma \in \Sigma) : \sigma = 0$ almost surely on $[0,x]\}$ and $\theta' = \inf\{x \in [0,1] | (\forall \sigma \in \Sigma) : \sigma = 1$ almost surely on $[x,1]\}$. Obviously, $\theta \leq \theta'$. Observe also that $\theta \leq BR_i(BR_i(\theta))$ if the disclosure game admits a unique BNE. Otherwise, the function that associates $x - BR_i(BR_i(x))$ to each x between 0 and θ is strictly positive at θ and non-positive at 0, and hence admits a zero by the intermediate values theorem. Let thus θ^* be an element of $[0,\theta)$ such that $\theta^* = BR_i(BR_i(\theta^*))$. Notice that the pair of strategies $(\sigma^{\theta^*}, \sigma^{BR_2(\theta^*)})$ then forms a BNE, which implies that $\sigma^{\theta^*} \in \Sigma$ and contradicts the definition of θ.

Any strategy in Σ for i's opponent has him withhold his information for almost every type between 0 and θ. The more his opponent reveals, the lower i's expected net gain, according to lemma 1. Hence if individual i wants to disclose his type when his opponent uses σ^θ, then a fortiori he wants to disclose it when his opponent plays some strategy in Σ (because there is more disclosure with σ^θ than with any strategy from Σ). This means that against any strategy in Σ, individual i's best response satisfies that he discloses his type whenever it is above $BR_i(\theta)$. Hence $\theta' \leq BR_i(\theta)$.

The third property in Lemma 1 implies that BR_i is non-increasing, and hence $BR_i(\theta') \geq BR_i(BR_i(\theta))$. In the same way we proved that $\theta' \leq BR_i(\theta)$, Lemma 1 and the definition of θ implies that $\theta \geq BR_i(\theta')$, and hence $\theta \geq BR_i(BR_i(\theta))$, by transitivity. Combining this with our earlier observation, we conclude that $\theta = BR_i(BR_i(\theta))$ and hence the pair of strategies $(\sigma^\theta, \sigma^{BR_2(\theta)})$ forms a BNE. Uniqueness of the BNE implies that this is in fact the symmetric

[2]Indeed, the set of strategies that survive the iterated elimination of strictly dominated strategies is the same for both individuals because the game is symmetric.

BNE. Hence we must also have that $\theta = BR_i(\theta)$, which implies that $\theta' = \theta$, and we are done proving that the unique symmetric BNE is also the unique profile of strategies that survive the iterated elimination of strictly dominated strategies when the disclosure game admits a unique BNE. ■

4 Step 1 in Proof of Proposition 7 (Sufficiency)

The proof is decomposed in two sub-steps.

Step 1.1 $\int_{y\in\tau^{-1}(\infty)} f(y)dy = 0.$

Proof: Suppose, to the contrary of what we want to prove, that $\int_{y\in\tau^{-1}(\infty)} f(y)dy > 0$. Let $x > 0$ be such that $\tau(x) = \infty$. The first individual's expected net gain from disclosing at a time t instead of ∞ is:

$$e^{-\delta t}x\int_{y\in\tau^{-1}(\infty)} f(y)dy + \int_{y\in\tau^{-1}(]t,\infty[)} (e^{-\delta t}x - e^{-\delta\tau(y)}(1-y))f(y)dy$$

$$+\int_{y\in\tau^{-1}(\{t\})} e^{-\delta t}(r_1(x,y) - (1-y))f(y)dy,$$

which is equal to $e^{-\delta t}$ times

$$x\int_{y\in\tau^{-1}(\infty)} f(y)dy + \int_{y\in\tau^{-1}(]t,\infty[)} (x - e^{-\delta(\tau(y)-t)}(1-y))f(y)dy$$

$$+\int_{y\in\tau^{-1}(\{t\})} (r_1(x,y) - (1-y))f(y)dy,$$

which is greater or equal to

$$x\int_{y\in\tau^{-1}(\infty)} f(y)dy - \int_{y\in\tau^{-1}(]t,\infty[)} f(y)dy,$$

since both x and $r_1(x,y)$ are non-negative, and both $1-y$ and $e^{-\delta(\tau(y)-t)}(1-y)$ are no larger than 1. The first term of this last expression is strictly positive, and independent of t, while the second can be made as small as needed by taking t large enough, as

$$\lim_{t\to\infty}\int_{y\in\tau^{-1}(]t,\infty[)} f(y)dy = 0,$$

by the measurability of τ. □

Step 1.2 *If $t \in]0,\infty[$, then $\int_{y\in\tau^{-1}(t)} f(y)dy = 0$.*

Proof: Let $\bar{x}$ be the supremum of $\tau^{-1}(t)$, and $\underline{x}$ be the infimum of $\tau^{-1}(t)$. For expositional convenience, we start by assuming that both the infimum and the supremum are reached in $\tau^{-1}(\infty)$, but we will show at the end of the proof how our argument extends to the more general case.

We start by assuming that $\underline{x} \leq 1-\bar{x}$. Hence $1-y \geq r_1(\underline{x},y)$, for all $y\in\tau^{-1}(t)$. In addition, $1-y > r_1(\underline{x},y)$ for each $y\in\tau^{-1}(t)$ such that $y<1/2$, as a consequence of the third regularity condition (Monotonicity), and the fact that $r_1(\underline{x},\underline{x}) = 1/2$. We now prove that $\int_{y\in\tau^{-1}(t)\cap[0,1/2[} f(y)dy = 0$. Otherwise, the previous reasoning implies that $\int_{y\in\tau^{-1}(t)}((1-y)-r_1(\underline{x},y))f(y)dy > 0$. Given that τ is a measurable function, we know that

$$\lim_{k\to\infty}\int_{y\in[0,1] \text{ s.t. } t<\tau(y)\leq t+\frac{1}{k}} f(y)dy = \int_{y\in[0,1] \text{ s.t. } t<\tau(y)\leq \lim_{k\to\infty} t+\frac{1}{k}} f(y)dy = 0,$$

and hence one can always find a k as large as necessary such that there is a very small probability for the other individual to speak in between t and $t+\frac{1}{k}$. The first individual's expected net gain of disclosing at $t+\frac{1}{k}$ instead of t when of type $\underline{x}$ is

$$\underline{x}(e^{-\delta(t+\frac{1}{k})}-e^{-\delta t})\int_{y\in\tau^{-1}(]t+\frac{1}{k},\infty])} f(y)dy + \int_{y\in\tau^{-1}(t+\frac{1}{k})}(e^{-\delta(t+\frac{1}{k})}r_1(\underline{x},y)-e^{-\delta t}\underline{x})f(y)dy$$

$$+\int_{y\in\tau^{-1}(]t,t+\frac{1}{k}[)}(e^{-\delta\tau(y)}(1-y)-e^{-\delta t}\underline{x})f(y)dy + \int_{y\in\tau^{-1}(t)} e^{-\delta t}((1-y)-r_1(\underline{x},y))f(y)dy,$$

which is larger or equal to $e^{-\delta t}$ times

$$\underline{x}(e^{-\delta/k-1})\int_{y\in\tau^{-1}(]t+\frac{1}{k},\infty])} f(y)dy - \underline{x}\int_{y\in\tau^{-1}(]t,t+\frac{1}{k}])} f(y)dy$$

$$+\int_{y\in\tau^{-1}(t)}((1-y)-r_1(\underline{x},y))f(y)dy,$$

as it is indeed easy to check that the integrand of the second and third terms from the previous expression are both larger or equal to $-\underline{x}e^{-\delta t}$. The first two terms of the last expression can be made as small as needed by choosing a k large enough, while the third one is strictly positive independently of k, and hence the possibility of a profitable deviation, which contradicts the fact that τ is part of a symmetric BNE. Hence we have proved, by contradiction, that $\int_{y\in\tau^{-1}(t)\cap[0,1/2[} f(y)dy = 0$, and hence that $\int_{y\in\tau^{-1}(t)} f(y)dy =$

$\int_{y\in\tau^{-1}(t)\cap[1/2,1]} f(y)dy$. If $\bar{x} \leq 1/2$, then we are done proving that $\int_{y\in\tau^{-1}(t)} f(y)dy = 0$. Let's thus assume that $\bar{x} > 1/2$.

Notice that $\bar{x} \geq r_1(\bar{x}, y)$, for each $y \in \tau^{-1}(t)$ such that $y \geq 1/2$. In fact, $\bar{x} > r_1(\bar{x}, y)$ for each $y \in \tau^{-1}(t)$ such that $y > 1/2$, as a consequence of condition (4) in the main paper, and the third regularity condition (Monotonicity). Hence $\int_{y\in\tau^{-1}(t)}(\bar{x} - r_1(\bar{x}, y))f(y)dy > 0$ if $\int_{y\in\tau^{-1}(t)\cap]1/2,1]} f(y)dy > 0$. In that case, one can construct a profitable deviation to a $t' < t$ for type $\bar{x}$ (similar argument to the one developed in the previous paragraph). To avoid this contradiction, on must accept that $\int_{y\in\tau^{-1}(t)\cap]1/2,1]} f(y)dy = 0$. Combined with the result of the previous paragraph, one concludes that $\int_{y\in\tau^{-1}(t)} f(y)dy = 0$, as desired.

A similar argument applies in the case where $\underline{x} \geq 1 - \bar{x}$, except that one must start to work with $\bar{x}$ to show that $\int_{y\in\tau^{-1}(t)\cap]1/2,1]} f(y)dy = 0$, and then work with $\underline{x}$ to conclude.

We now consider the case where $\underline{x}$ and $\bar{x}$ do not necessarily belong to $\tau^{-1}(t)$. Again, we provide the argument only for the case were $\underline{x} \leq 1 - \bar{x}$, a similar argument applying if the inequality is reversed. Let $(\underline{x}_n)_{n\in\mathbb{N}}$ be a decreasing sequence in $\tau^{-1}(t)$ that converges to $\underline{x}$, and let $(\bar{x}_n)_{n\in\mathbb{N}}$ be an increasing sequence in $\tau^{-1}(t)$ that converges to $\bar{x}$ such that $\underline{x}_n \leq 1 - \bar{x}_n$, for each n. For notational simplicity, let α_n be the following real number:

$$\alpha_n := \int_{y\in\tau^{-1}(t)\cap[\underline{x}_n,\bar{x}_n]} ((1-y) - r_1(\underline{x}_n, y))f(y)dy,$$

for each $n \in \mathbb{N}$. Notice first that these numbers are non-decreasing in n. Indeed, consider $m < n$. We have:

$$\begin{aligned}\alpha_n = &\int_{y\in\tau^{-1}(t)\cap[\underline{x}_n,\underline{x}_m]} ((1-y) - r_1(\underline{x}_n, y))f(y)dy \\ &+ \int_{y\in\tau^{-1}(t)\cap[\underline{x}_m,\bar{x}_m]} ((1-y) - r_1(\underline{x}_n, y))f(y)dy \\ &+ \int_{y\in\tau^{-1}(t)\cap[\bar{x}_m,\bar{x}_n]} ((1-y) - r_1(\underline{x}_n, y))f(y)dy.\end{aligned}$$

Since $\underline{x}_n \leq 1 - \bar{x}_n$, we must have $r_1(\underline{x}_n, y) \leq 1 - y$, for each $y \in [\underline{x}_n, \bar{x}_n]$, and hence the first and the third terms must be non-negative. The third regularity condition also implies that the second term is larger or equal to α_m, since $\underline{x}_m \geq \underline{x}_n$, and hence $\alpha_n \geq \alpha_m$, as desired.

We now show that $\int_{y\in\tau^{-1}(t)\cap[0,1/2]} f(y)dy = 0$. Otherwise, there exists N such that $\int_{y\in\tau^{-1}(t)\cap[0,1/2]\cap[\underline{x}_n,\bar{x}_n]} f(y)dy > 0$, for each $n \geq N$. The reasoning that we did at the beginning of the proof when the infimum and the supremum are reached implies that $\alpha_n > 0$, for each $n \geq N$, and in particular $\alpha_N > 0$. Notice that

$$\int_{y\in\tau^{-1}(t)} ((1-y)-r_1(\underline{x}_n, y))f(y)dy = \alpha_n + \int_{y\in\tau^{-1}(t)\setminus[\underline{x}_n,\bar{x}_n]} ((1-y)-r_1(\underline{x}_n, y))f(y)dy,$$

for each $n \geq N$. The first term is larger or equal to α_N, which is strictly larger than 0 and independent of n, while the second term converges towards zero as n increases, since the integrand is bounded and $\int_{y\in\tau^{-1}(t)\setminus[\underline{x}_n,\bar{x}_n]} f(y)dy$ converges towards zero, and we are done proving that the expression on the left-hand side must be strictly positive for n large enough. As before, this implies that the first individual of type $\underline{x}_n$ prefers to disclose his type slightly later than at t, thereby contradicting the definition of a BNE. It must thus be the case that $\int_{y\in\tau^{-1}(t)\cap[0,1/2]} f(y)dy = 0$, as desired.

Adapting the argument to show that $\int_{y\in\tau^{-1}(t)\cap[1/2,1]} f(y)dy = 0$ when the infimum and the supremum are not reached, and thereby conclude the proof, is similar and left to the reader. ■

5 Proof of Proposition 11 (Necessity)

We prove that the strategy $\mathbf{t}^*$ is indeed part of a symmetric BNE.

We start by showing that reporting at $\mathbf{t}^*(x)$ is optimal, for any $x \in [0, 1/2[$. Consider first the possibility of disclosing at positive times. The function $\mathbf{t}^*$ being invertible on $[0, 1/2[$, we can identify any positive time with the type speaking at that time. The expected utility from disclosing at $\mathbf{t}^*(z)$ when of type x is equal to

$$U(z|x) := xF(z)e^{-\delta\mathbf{t}^*(z)} + \int_{y=z}^{1-x} (1-y)e^{-\delta\mathbf{t}^*(y)} f(y)dy + \int_{y=1-x}^{1} r_1(x, y)e^{-\delta\mathbf{t}^*(y)} f(y)dy,$$

for each $z \in [0, 1/2[$. It is easy to check that this expression is differentiable, and has the same derivative as the similar expression in the proof of Proposition 7 (because the third term does not depend on z), i.e.

$$\frac{(1-z)}{z} f(z)(x-z)e^{-\delta\mathbf{t}^*(z)}.$$

We see that the first order condition is satisfied at $z = x$, and that the derivative is positive when $z < x$ and negative when $x < z$. Hence there is no profitable deviation to a positive time different from $\mathbf{t}^*(x)$, when of type x.

Deviating to report at zero is not profitable either, as the expected payoff in that case is

$$\frac{x}{2} + \int_{y=1/2}^{1} r_1(x,y) f(y) dy$$

which is equal to

$$U(1/2|x) + \int_{y=1/2}^{1-x} (r_1(x,y) - (1-y)) f(y) dy,$$

and the second term of this expression is non-positive, as $y \leq 1-x$ implies that $x \leq 1-y$ and hence $r_1(x,y) \leq 1-y$.

Consider now a type $x \in [1/2, 1]$. The expected utility of disclosing at a positive time t, corresponding to a $z < 1/2$, is equal to

$$xF(z)e^{-\delta \mathfrak{t}^*(z)} + \int_{y=z}^{1-x} (1-y) e^{-\delta \mathfrak{t}^*(y)} f(y) dy + \int_{y=1-x}^{1} r_1(x,y) e^{-\delta \mathfrak{t}^*(y)} f(y) dy,$$

if $z \leq 1-x$, and to

$$xF(z)e^{-\delta \mathfrak{t}^*(z)} + \int_{y=z}^{1} r_1(x,y) e^{-\delta \mathfrak{t}^*(y)} f(y) dy,$$

if $z \geq 1-x$. The expression when $z \leq 1-x$ is non-decreasing in z, as was $U(z|x)$ when x was smaller than $1/2$. The expression when $z \geq 1-x$ is also non-decreasing because its derivative is equal to

$$(\frac{1-z}{z} x - r_1(x,z)) f(z) e^{-\delta \mathfrak{t}^*(z)}.$$

Notice that $z \geq 1-x$ implies $x \geq 1-z$ and hence $r_1(x,z) \leq x$. On the other hand, $z \leq 1/2$ implies that $(1-z)/z \geq 1$, and hence $r_1(x,z) \leq (x(1-z))/z$, which implies that the last derivative in non-negative, as desired. The expected utility of disclosing at a positive t is thus no larger than when taking the limit of that expected utility when z tends to $1/2$, i.e. $\frac{x}{2} + \int_{y=1/2}^{1} r_1(x,y) f(y) dy$. But this is exactly the expected utility the individual gets by disclosing at zero, which shows that there are no profitable deviations when $x \in [1/2, 1]$ either. ■

6 Proof of Proposition 11 (Sufficiency)

Let r be a regular compromise rule, and let $\mathfrak{t}$ be a strategy that is part of a refined symmetric BNE in the dynamic game with an opportunity to react. We have to show that $\mathfrak{t} = \mathfrak{t}^*$. We proceed in various steps.

Step 1 $\int_{y\in \mathfrak{t}^{-1}(t)\cap[0,1/2]} f(y)dy = 0$*, for all* $t \in \mathbb{R}_+$*.*

Proof: The first individual's expected net gain of disclosing at $t' > t$ instead of t, when of type x, is equal to

$$\int_{y\in \mathfrak{t}^{-1}(]t',\infty])} \min\{x, r_1(x,y)\}(e^{-\delta t'} - e^{-\delta t})f(y)dy$$

$$+\int_{y\in \mathfrak{t}^{-1}(t')} (r_1(x,y)e^{-\delta t'} - \min\{x, r_1(x,y)\}e^{-\delta t})f(y)dy$$

$$+\int_{y\in \mathfrak{t}^{-1}(]t,t'[)} (\max\{1-y, r_1(x,y)\}e^{-\delta \mathfrak{t}(y)} - \min\{x, r_1(x,y)\}e^{-\delta t})f(y)dy$$

$$+\int_{y\in \mathfrak{t}^{-1}(t)} (\max\{1-y, r_1(x,y)\} - r_1(x,y))e^{-\delta t}f(y)dy,$$

which is larger or equal to

$$\int_{y\in \mathfrak{t}^{-1}(]t',\infty])} \min\{x, r_1(x,y)\}(e^{-\delta t'} - e^{-\delta t})f(y)dy - \int_{y\in \mathfrak{t}^{-1}(]t,t'])} \min\{x, r_1(x,y)\}e^{-\delta t}f(y)dy$$

$$+\int_{y\in \mathfrak{t}^{-1}(t)} (\max\{1-y, r_1(x,y)\} - r_1(x,y))e^{-\delta t}f(y)dy,$$

since the integrand of the second and third terms are both larger or equal to $-\min\{x, r_1(x,y)\}$.

Suppose, to the contrary of what we want to prove, that $\int_{y\in \mathfrak{t}^{-1}(t)\cap[0,1/2]} f(y)dy > 0$, for some $t \geq 0$. Let's focus on one of the types x that discloses at t, and that is small enough so that $\int_{y\in \mathfrak{t}^{-1}(t)\cap[x,1/2]} f(y)dy > 0$. Notice that $\max\{1-y, r_1(x,y)\} \geq r_1(x,y)$, for any $y \in [0,1]$, and that $\max\{1-y, r_1(x,y)\} > r_1(x,y)$, for any $y \in [x, 1/2[$. Indeed, the second regularity condition implies that $r_1(x,x) = 1/2$, and the third regularity condition implies that $r_1(x,y) \leq 1/2 < 1-y$, for all such y's. Hence the third term in the lower bound on the first individual's expected net gain of disclosing at t' instead of t is strictly positive, and independent of t'. The first two terms, on the other hand, can be made as small as needed by choosing t' close enough to t (similar to argument developed in the second paragraph of the proof of Step 1.2 in Section 3 of this Supplementary Appendix), thereby leading to a contradiction of the optimality of $\mathfrak{t}$. $\square$

Step 2 *Let* $x, x' \in [0,1]$ *be such that* $x' < 1/2 < x$*. If* $\int_{y\in \mathfrak{t}^{-1}([\mathfrak{t}(x),\infty])} f(y)dy > 0$*, then* $\mathfrak{t}(x') \geq \mathfrak{t}(x)$*.*

Proof: Let $t = \mathfrak{t}(x)$ and $t' = \mathfrak{t}(x')$. Suppose, to the contrary of what we want to prove, that $t > t'$. The first individual's expected net gain of disclosing at t instead of t', when of type x, is equal to

$$\int_{y \in \mathfrak{t}^{-1}(]t,\infty])} \min\{x, r_1(x,y)\}(e^{-\delta t} - e^{-\delta t'})f(y)dy$$

$$+ \int_{y \in \mathfrak{t}^{-1}(t)} (r_1(x,y)e^{-\delta t} - \min\{x, r_1(x,y)\}e^{-\delta t'})f(y)dy$$

$$+ \int_{y \in \mathfrak{t}^{-1}(]t',t[)} (\max\{1-y, r_1(x,y)\}e^{-\delta \mathfrak{t}(y)} - \min\{x, r_1(x,y)\}e^{-\delta t'})f(y)dy$$

$$+ \int_{y \in \mathfrak{t}^{-1}(t')} (\max\{1-y, r_1(x,y)\} - r_1(x,y))e^{-\delta t'} f(y)dy.$$

We now prove that this expected net gain does not decrease when replacing x by x'. The third regularity condition implies that $\min\{x, r_1(x,y)\}$ is non-decreasing in x, for all $y \in [0,1]$. Hence $\min\{x, r_1(x,y)\}(e^{-\delta t} - e^{-\delta t'})$ is non-increasing in x, as $t > t'$. If $t = \infty$, then $r_1(x,y)e^{-\delta t} - \min\{x, r_1(x,y)\}e^{-\delta t'} = -\min\{x, r_1(x,y)\}e^{-\delta t'}$, which again is non-increasing in x, independently of y. If t is finite, then the integral in the second term is equal to the integral when $y \geq 1/2$, by Step 1. The integrand in that case is equal to $r_1(x,y)(e^{-\delta t} - e^{-\delta t'})$. The integrand has the same functional form when x is replaced by x', for all y's such that $1 - y \leq x'$, which is thus no smaller than what it was with x, by the third regularity condition. Consider now some y such that $1 - y \in]x', 1/2[$. We have:

$$r_1(x,y)(e^{-\delta t} - e^{-\delta t'}) \leq r_1(1-y,y)(e^{-\delta t} - e^{-\delta t'}) = \min\{r_1(1-y,y), 1-y\}(e^{-\delta t} - e^{-\delta t'})$$

$$\leq \min\{r_1(x',y), x'\}(e^{-\delta t} - e^{-\delta t'}) \leq r_1(x',y)e^{-\delta t} - \min\{r_1(x',y), x'\}e^{-\delta t'},$$

where the two first inequalities follow from the third regularity condition, since $x' < 1 - y < x$, and the equality follows from the fact that $r_1(1-y,y) = 1-y$. Let's consider now the integrand of the third term. First, if $1 - y > x$, then it is equal to $(1-y)e^{-\delta \mathfrak{t}(y)} - xe^{-\delta t'}$. Then $1 - y > x'$ a fortiori, and therefore the integrand is equal to $(1-y)e^{-\delta \mathfrak{t}(y)} - x'e^{-\delta t'}$ when x is replaced by x', which is strictly greater than the previous expression. If $1 - y < x'$, then the integrand for x' is equal to $r_1(x',y)(e^{-\delta \mathfrak{t}(y)} - e^{-\delta t'})$, which is no smaller than the integrand for x, which is equal to $r_1(x,y)(e^{-\delta \mathfrak{t}(y)} - e^{-\delta t'})$. A similar comparison holds when $x' < 1 - y < x$:

$$\max\{1-y, r_1(x,y)\}e^{-\delta \mathfrak{t}(y)} - \min\{x, r_1(x,y)\}e^{-\delta t'} = r_1(x,y)(e^{-\delta \mathfrak{t}(y)} - e^{-\delta t'})$$

$\leq (1-y)(e^{-\delta \mathfrak{t}(y)} - e^{-\delta t'}) \leq \max\{1-y, r_1(x',y)\}e^{-\delta t} - \min\{x', r_1(x',y)\}e^{-\delta t'}$.

Finally, Step 1 implies that we can restrict attention to $y \geq 1/2$ in the fourth term. In that case, the integrand is equal to zero when of type x, while the integrand for x' is non-negative.

Given that $\int_{y\in \mathfrak{t}^{-1}([t,\infty])} f(y)dy > 0$, there must be a positive probability that the second individual discloses y for which $1-y > x$ strictly after t' and strictly before t. Notice indeed that all the terms associated to other y's in the first individual's expected net gain of disclosing at t instead of t', when of type x, are non-positive, and in fact must sum up to a strictly negative number when the second individual discloses a collective action with positive probability after t. Remember our reasoning from the previous paragraph that the integrand involving y's such that $1-y > x$, and that are disclosed strictly after t' and strictly before t, are strictly increasing when replacing x by x'. If $\mathfrak{t}$ is part of a symmetric BNE, then it must be that the first individual's expected net gain of disclosing his collective action at t instead of t' is non-negative when of type x, but our reasoning also shows that the same expected net gain is strictly larger for x' if $t > t'$, thereby contradicting the optimality of $\mathfrak{t}$. We have thus shown that $t \leq t'$, as desired. $\square$

Step 3 $\mathfrak{t}(x) = 0$*, for almost all* $x \in]1/2, 1]$*, i.e.* $\int_{x\in]1/2,1]\cap \mathfrak{t}^{-1}(]0,\infty])} f(x)dx = 0$.

Proof: Let X be the set of x's in $]1/2, 1]$ such that $\int_{y\in \mathfrak{t}^{-1}([\mathfrak{t}(x),\infty])} f(y)dy > 0$, and $\bar{X}$ be its complement in $]1/2, 1]$. Let also t be the infimum of $\mathfrak{t}(x)$ when x varies in $\bar{X}$, and let $(t_k)_{k\in\mathbb{N}}$ be a decreasing sequence of non-negative real number such that $(t_k)_{k\in\mathbb{N}}$ converges to t, and $t_k = \mathfrak{t}(x_k)$ for some $x_k \in \bar{X}$, for each $k \in \mathbb{N}$. We have:

$$\int_{x\in\bar{X}} f(x)dx \leq \int_{y\in \mathfrak{t}^{-1}([t,\infty])} f(y)dy = \lim_{k\to\infty} \int_{y\in \mathfrak{t}^{-1}([t_k,\infty])} f(y)dy = 0.$$

We will now show that $\mathfrak{t}(x) = 0$, for all $x \in X$. This will allow us to conclude the proof, since the probability of an individual not disclosing his collective action at $t = 0$ when of a type $x \in]1/2, 1]$ will then be known to be no larger than the probability of $\bar{X}$, which we have just shown is null.

Let thus $x \in]1/2, 1]$ be such that $\int_{y\in \mathfrak{t}^{-1}([t(x),\infty])} f(y)dy > 0$. Suppose, to the contrary of what we want to prove that $t = \mathfrak{t}(x) > 0$. The first individual's expected net gain of disclosing at 0 instead of t is equal to

$$\int_{y\in \mathfrak{t}^{-1}(]t,\infty])} \min\{x, r_1(x,y)\}(1-e^{-\delta t})f(y)dy$$

$$+\int_{y\in\mathfrak{t}^{-1}(t)}(\min\{x,r_1(x,y)\}-r_1(x,y)e^{-\delta t})f(y)dy$$

$$+\int_{y\in\mathfrak{t}^{-1}(]0,t[)}(\min\{x,r_1(x,y)\}-\max\{1-y,r_1(x,y)\}e^{-\delta\mathfrak{t}(y)})f(y)dy$$

$$+\int_{y\in\mathfrak{t}^{-1}(0)}(r_1(x,y)-\max\{1-y,r_1(x,y)\})f(y)dy.$$

The integrand in the first term is clearly strictly positive. The integral in the second term can be restricted to those y's that are no smaller than $1/2$, by Step 1, and the integrand is equal to $r_1(x,y)(1-e^{-\delta t})$. Again, this is strictly positive. We know from Step 2 that y must be at least $1/2$ to be disclosed strictly before t. Hence $1-y<x$ for all such y's, and the third integrand is equal to $r_1(x,y)(1-e^{-\delta\mathfrak{t}(y)})$, which is strictly positive when $\mathfrak{t}(y)>0$, while the integrand in the fourth term is null. Given that there is a positive probability that the other individual discloses his collective action at or after t, one concludes that the first individual's expected net gain of disclosing at 0 instead of t is strictly positive, which contradicts the optimality of $\mathfrak{t}$. Hence $\mathfrak{t}(x)=0$, and we are done with the proof of this step. □

Step 4 $\mathfrak{t}$ *is strictly decreasing on* $[0,1/2[$.

Proof: Consider $x'<x<1/2$, and let $t=\mathfrak{t}(x)$ and $t'=\mathfrak{t}(x')$. Let's start by assuming that $\int_{y\in\mathfrak{t}^{-1}([\mathfrak{t}(x),\infty])}f(y)dy>0$. It is straightforward to check that the proof of Step 2 goes through in this case as well, after noticing that the second term in the expected net gain of disclosing at t instead of t' is null when $t>t'$, as the probability of an individual disclosing at a strictly positive time is null thanks to Steps 1 and 3. Hence $\mathfrak{t}(x')\geq\mathfrak{t}(x)$.

We may assume that $\mathfrak{t}(x)>0$, as otherwise almost all types between x and $1/2$ disclose at 0, contradicting Step 1. We know from the previous paragraph that $\mathfrak{t}(x'')\leq\mathfrak{t}(x)$, for all $x''\in]x',x[$. Step 1 implies that there exists $x''\in]x',x[$ such that $\mathfrak{t}(x'')>\mathfrak{t}(x)$. The reasoning from the previous paragraph implies that $\mathfrak{t}(x')\geq\mathfrak{t}(x'')$, and hence $\mathfrak{t}(x')>\mathfrak{t}(x)$.

We have thus established the desired property, but under the additional assumption that $\int_{y\in\mathfrak{t}^{-1}([t(x),\infty])}f(y)dy>0$. We now show that this inequality must in fact hold for any $x\in]0,1/2[$, thereby proving the result by applying our previous arguments to x's that are as close to $1/2$ as needed. Let x^* be the supremum of the x's in $[0,1/2[$ for which there is a strictly positive probability of disclosure on or after $\mathfrak{t}(x)$. We thus have to show that $x^*=1/2$. Suppose on the contrary that $x^*<1/2$. Let then $t^*=\inf_{y\in]x^*,1/2[}\mathfrak{t}(y)$, and let $(x_k)_{k\in\mathbb{N}}$

be a sequence in $]x^*, 1/2[$ such that $(\mathfrak{t}(x_k))_{k\in\mathbb{N}}$ decreases towards t^*, as k tends to infinity. Since $\mathfrak{t}$ is measurable, we have:

$$\lim_{k\to\infty}\int_{y\in\mathfrak{t}^{-1}([\mathfrak{t}(x_k),\infty])} f(y)dy = \int_{y\in\mathfrak{t}^{-1}([\lim_{k\to\infty}\mathfrak{t}(x_k),\infty])} f(y)dy = \int_{y\in\mathfrak{t}^{-1}([t^*,\infty])} f(y)dy.$$

Notice that the right-most expression must be strictly positive, since $]x^*, 1/2[\subseteq \mathfrak{t}^{-1}([t^*,\infty])$, by definition of t^*. Hence there exists $K\in\mathbb{N}$ such that $\int_{y\in[0,1]\text{ s.t. }\mathfrak{t}(y)\geq\mathfrak{t}(x_k)} f(y)dy > 0$, for all $k\geq K$, leading to the desired contradiction, given the definition of x^*. $\square$

Step 5 $\mathfrak{t} = \mathfrak{t}^*$.

Proof: We start by strengthening the result from Step 3, by showing that $\mathfrak{t}(x) = 0$, for all $x > 1/2$. Suppose, to the contrary of what we want to prove, that $\mathfrak{t}(x) > 0$, for some $x > 1/2$. Let us compute type x's expected net gain of disclosing at $\mathfrak{t}(x)$ instead of 0. This expression is the same as the one written in the proof of Step 2, if one takes $t = \mathfrak{t}(x)$ and $t' = 0$. Notice also that the second term in the formula is null, since almost all types above 1/2 disclose at zero (cf. Step 3), and the revelation strategy followed by types smaller than 1/2 is strictly decreasing (cf. Step 4). The fourth term is zero as well, because $y \geq 1/2$ if disclosed at zero (cf. Step 4), and $\max\{1-y, r_1(x,y)\} = r_1(x,y)$, for all such y's. Hence the expected net gain can be rewritten as follows:

$$\int_{y\in\mathfrak{t}^{-1}([t,\infty])} \min\{x, r_1(x,y)\}(e^{-\delta t}-1)f(y)dy + \int_{y\in\mathfrak{t}^{-1}([0,t]),\ y\geq 1-x} r_1(x,y)(e^{-\delta\mathfrak{t}(y)}-1)f(y)dy$$

$$+\int_{y\in\mathfrak{t}^{-1}([0,t]),\ y\leq 1-x} ((1-y)e^{-\delta\mathfrak{t}(y)} - x)f(y)dy.$$

Notice that this expression is strictly negative if $\int_{y\in\mathfrak{t}^{-1}([0,t]),\ y\leq 1-x} f(y)dy = 0$, which would contradict the optimality of disclosing at $t = \mathfrak{t}(x) > 0$ when of type x. Consider now the expected net gain for a type $x' \in]1/2, x[$ to disclose at t instead of 0. A simple rearrangement of terms in the integrals implies that it is equal to

$$\int_{y\in\mathfrak{t}^{-1}([t,\infty])} \min\{x', r_1(x',y)\}(e^{-\delta t}-1)f(y)dy + \int_{y\in\mathfrak{t}^{-1}([0,t]),\ y\geq 1-x} r_1(x',y)(e^{-\delta\mathfrak{t}(y)}-1)f(y)dy$$

$$+\int_{y\in\mathfrak{t}^{-1}([0,t]),\ y\leq 1-x} ((1-y)e^{-\delta\mathfrak{t}(y)} - x')f(y)dy$$

$$+\int_{y\in \mathfrak{t}^{-1}([0,t]),\ 1-x\leq y\leq 1-x'}[((1-y)-r_1(x',y))e^{-\delta \mathfrak{t}(y)}+(r_1(x',y)-x')]f(y)dy.$$

The first two terms are no smaller than their counterpart with x instead of x'. The third term, on the other hand, is strictly larger than its counterpart, since $\int_{y\in \mathfrak{t}^{-1}([0,t]),\ y\leq 1-x} f(y)dy > 0$. The fourth term, finally, is non-negative since $y \leq 1-x'$ implies $x' \leq r_1(x',y) \leq 1-y$. Type x's expected net gain of disclosing at t instead of 0 being non-negative, it must now be strictly positive for type x'. Hence all the types in $]1/2, x']$ would disclose after 0, thereby contradicting Step 3. This establishes that $\mathfrak{t}(x) = \mathfrak{t}^*(x)$, for all $x > 1/2$.

Next, one can follows the arguments in the proofs of Steps 4 and 6 in the proof of Proposition 7 (see Appendix in main paper) to show that $\mathfrak{t}$ is continuous of $]0, 1/2[$, that $\lim_{x\to 1/2_-} \mathfrak{t}(x) = 0$, and that $\mathfrak{t}$ is differentiable on $]0, 1/2[$ with

$$\mathfrak{t}'(x) = \frac{(1-2x)f(x)}{\delta x F(x)},$$

for each $x \in]0, 1/2[$. One can then follow the argument from the proof of Step 7 in the proof of Proposition 7 (see Appendix in main paper) to show that $\mathfrak{t} = \mathfrak{t}^*$. ■

7 r_{CF}, r_{KS}, r_N and r^* Are Regular

By definition, all four compromise rules are anonylous and ex-post efficient. It remains to verify that they are also monotone. By definition, the coin-flip rule is monotone regardless of whether g is convex or not. To show that the Nash solution is monotone, let $(x, g(x))$ and $(y, g(y))$ be two payoff pairs on the utility frontier $u_2 = g(u_1)$ such that $y > x$. The line connecting these two points is given by

$$u_2 = g(y) + \alpha(x, y) \cdot (y - u_1)$$

where

$$\alpha(x, y) \equiv \frac{g(x) - g(y)}{y - x}$$

Let $r_i^N(x, y)$ be individual i's payoff at the Nash solution associated with $(x, g(x))$ and $(y, g(y))$. The first individual's payoff under the Nash solution is as close as possible to half the intercept of the line going through $((x, g(x))$ and $(y, g(y))$, and hence

$$r_1^N(x,y) = \begin{cases} \phi(x,y) & if \quad x < \phi(x,y) < y \\ x & if \quad \phi(x,y) \leq x \\ y & if \quad \phi(x,y) \geq y, \end{cases}$$

where

$$\phi(x,y) \equiv \frac{g(y)}{2\alpha(x,y)} + \frac{y}{2}.$$

Consider first a change from $x = z$ to $x = z'$ such that $y > z' > z$. We need to show that $r_1^N(z',y) \geq r_1^N(z,y)$ and $r_2^N(z',y) \leq r_2^N(z,y)$. Note that because $\alpha(z',y) < \alpha(z,y)$ we have that $\phi(z',y) > \phi(z,y)$. A priori there are nine cases to consider, with z and z' falling in the three different areas that define r_1^N. It is straightforward to show that monotonicity does occur, or that the combination of conditions are impossible, in all except perhaps the following two cases. If z falls in the first region ($x < \phi(z,y) < y$), while z' falls in the second region ($\phi(z',y) \leq z'$), then $r_1^N(z,y) = \phi(z,y) \leq \phi(z',y) \leq z' = r_1^N(z',y)$, and we are done proving monotonicity in that case. Also, it is impossible for z to fall in the third area, and for z' to fall in in the first or second area, since this would lead to the contradiction $y \leq \phi(z,y) < \phi(z',y) < y$. It follows that $r_1^N(z',y) \geq r_1^N(z,y)$. An analogous argument shows that $r_2^N(z',y) \leq r_2^N(z,y)$, and that monotonicity is satisfied when y changes from $y = z$ to $y = z'$ such that $z' > z > x$.

As for the Kalai-Smorodinsky solution, let $(x, g(x))$ and $(y, g(y))$ be two points on the frontier satisfying $y > x$ (and hence, $g(x) > g(y)$). The KS solution to $(x, g(x))$ and $(y, g(y))$ is given by the intersection of the line connecting the two points with the ray going from the origin to the "utopia" point $(y, g(x))$.

Suppose we increase y to y'. By the definition of KS, it is clear that the expected payoff of the first individual assigned by KS will increase. It is not clear what happens to the second individual's expected payoff. Let u_2 be the second individual's expected payoff in the KS solution to $(x, g(x))$ and $(y, g(y))$. Let u_2' be the second individual's expected payoff at the solution assigned to $(x, g(x))$ and $(y', g(y'))$. We want to show that $u_2 > u_2'$.[3]

The KS solution to $(x, g(x))$ and $(y, g(y))$ is given by the equation

$$\frac{y}{g(x)} u_2 = x + [\frac{y-x}{g(x)-g(y)}][g(x) - u_2]$$

Let

$$\delta \equiv \frac{y}{g(x)}$$

(the inverse of the slope of the ray) and

$$\mu \equiv \frac{y-x}{g(x)-g(y)}$$

[3] Renaming variables implies that the subsequent reasoning also applies when decreasing y to y', as long as y' remains above x. In that case, $u_2' > u_2$, as needed for monotonicity.

(the inverse of the absolute value of the slope of the line connecting the two points on the frontier). In a similar way, define

$$\delta' \equiv \frac{y'}{g(x)}$$

and

$$\mu' \equiv \frac{y' - x}{g(x) - g(y')}$$

We can therefore solve for u_2 and u_2' :

$$u_2 = \frac{x + \mu g(x)}{\delta + \mu}$$

and

$$u_2' = \frac{x + \mu' g(x)}{\delta' + \mu'}$$

Assuming $y' > y$, we want to show that $u_2 > u_2'$, or

$$\frac{x + \mu g(x)}{\delta + \mu} > \frac{x + \mu' g(x)}{\delta' + \mu'}$$

which is equivalent to (since the denominators are positive)

$$x(\delta' + \mu' - \delta - \mu) + g(x)(\mu\delta' - \mu'\delta) > 0$$

Since $g(x) = y'/\delta' = y/\delta$, this is equivalent to

$$x(\delta' + \mu' - \delta - \mu) + y'\mu - y\mu' > 0$$

which may be rewritten as

$$\mu(y' - x) - \mu'(y - x) + x(\delta' - \delta) > 0$$

Plugging in the expressions for $(\mu, \mu', \delta, \delta')$ gives

$$\frac{(y - x)(y' - x)}{g(x) - g(y)} - \frac{(y' - x)(y - x)}{g(x) - g(y')} + \frac{x(y' - y)}{g(x)} > 0$$

Placing the first two terms under the same denominator, it thus amounts to show

$$\frac{(y - x)(y' - x)[g(y) - g(y')]}{[g(x) - g(y)][g(x) - g(y')]} + \frac{x(y' - y)}{g(x)} > 0$$

The inequality indeed holds, as $y' > y > x$ and $g(x) > g(y) > g(y')$.

The fact that the first individual's payoff increases (decreases) and the second individual's payoff decreases (increases) when increasing x to x' whenever both x and x' fall above y, follows from the previous argument, after observing that the Kalai-Smorodinsky solution is anonymous.

We conclude the proof by checking that r^* is monotone. This follows at once from the definition in the following cases:
(i) Start from two points on the same side of the 45 degree line $u_2 = u_1$ and change only one of the points such that both still remain on the same side of $u_2 = u_1$.
(ii) Start from $(x, g(x))$ and $(z, g(z))$ such that $g(x) > x$, $g(z) < z$ and $g(x) \geq z$. Fix $(x, f(x))$ and change $(z, f(z))$ into $(z', f(z'))$ such that it is still the case that $g(x) \geq \max\{z', g(z')\}$.

Monotonicity is more difficult to show in the last remaining case (all other cases follow by symmetry): starts from $(x, g(x))$ and $(z, g(z))$ such that $g(x) > x$, $g(z) < z$, $g(x) > z$ and $g(z) > x$, then change $(x, g(x))$ into $(x', g(x'))$ such that $g(x') > z$.

We will prove monotonicity by checking the sign of the derivative of r_1^* with respect to its first component in that last region. It is helpful to do the following change of variable. For each $(x, g(x))$ falling in that last region, let α be the absolute value of the slope of the line joining $(z, g(z))$ to $(x, g(x))$. Vice versa, each $\alpha > 1$ determines a unique $(x, g(x))$ that falls in that region (at the intersection of X and the line of slope $-\alpha$ that goes through $(z, g(z))$). Let $\delta = x + g(x)$ (note that this is the utilitarian surplus). Then, for each $\alpha > 1$, we have:

$$\delta(\alpha)/2 = g(z) + \alpha(z - r_1^*(x(\alpha), g(z))),$$

or

$$r_1^*(x(\alpha), g(z)) = z - \frac{\delta(\alpha) - 2g(z)}{2\alpha}.$$

Let now ϵ be any small strictly positive number. We have:

$$\frac{r_1^*(x(\alpha+\epsilon), g(z)) - r_1^*(x(\alpha), g(z))}{\epsilon} = \frac{\delta(\alpha)\alpha + \delta(\alpha)\epsilon - 2g(z)\epsilon - \alpha\delta(\alpha+\epsilon)}{2\alpha(\alpha+\epsilon)\epsilon}.$$

Taking the limit as epsilon tends to zero, this expression is equal to

$$-\frac{\delta'(\alpha)}{2\alpha} + \frac{\delta(\alpha) - 2g(z)}{2\alpha^2}$$

(δ is differentiable because g is). Notice that $\delta(\alpha + \epsilon)$ is larger than the sum of the components of the vector at the intersection of this new line (going through $(z, g(z))$ and with angle $-\alpha - \epsilon$) and the vertical line going through $(x, g(x))$. This is so because the intersection of the new line with the utility frontier falls on the left of x, and the slope $\alpha + \epsilon$ is larger than 1 (i.e. any decrease in the first component is more than matched by an increase in the second component). The sum of the components of the vector associated to the new line is $x + g(x) + (z - x)\epsilon$. Therefore,

$$\delta'(\alpha) = \lim_{\epsilon \to 0} \frac{\delta(\alpha + \epsilon) - \delta(\alpha)}{\epsilon} \geq \lim_{\epsilon \to 0} \frac{x + g(x) + (z - x)\epsilon - x - g(x)}{\epsilon} = z - x.$$

Hence

$$\frac{dr_1^*(x(\alpha), g(z))}{d\alpha} \leq \frac{-\alpha(z - x) + \delta(\alpha) - 2g(z)}{2\alpha^2} = \frac{x - g(z)}{2\alpha^2} \leq 0,$$

where the equality follows from the fact that $\alpha(z, x) = g(x) - g(z)$ and $\delta(\alpha) = x + g(x)$, and the last inequality follows from the fact that $x \leq g(z)$ (because $g(x) \geq z$, $g(z) < z$ and $g(x) > x$). Finally, $dx/d\alpha$ being strictly negative, it must be that r_1^* varies monotonically with x, as desired. ■

8 $r_{KS} \succ r_{CF}$

Let $x \leq 1/2$ and $y \geq x$. We need to prove that the first individual's payoff under the Kalai-Smorodinsky solution is larger than his payoff under the coin-flip rule when he reports x and his opponent reports y. Consider first the case where $y \leq 1 - x$, for which the relevant inequality to check is

$$\frac{1 - y}{(1 - y) + (1 - x)} \geq \frac{x + (1 - y)}{2}.$$

Simple algebra shows that this inequality is equivalent to $0 \geq x(1-x) - y(1-y)$, which indeed holds true since the function $h(z) = z(1-z)$ is symmetric around 1/2, increasing before 1/2 and decreasing after 1/2. Similarly, the relevant inequality to check when $y \geq 1 - x$ is

$$\frac{x}{x + y} \geq \frac{x + (1 - y)}{2}.$$

Simple algebra shows that this inequality is equivalent to $0 \geq -x(1 - x) + y(1 - y)$, which again holds true because of the properties of the function.

9 Inefficiency with Proportional Solutions

We focus on the symmetric proportional rule applied to the static disclosure game in our benchmark model. The rule picks the utility pair that is as far as possible from $(0, 0)$ along the 45-degree line. Thus, following the disclosure stage, the symmetric proportional rule picks the utility pair $(1/2, 1/2)$ whenever it is feasible (i.e., whenever the two disclosed options fall on opposite sides of the 45 degree line), and picks the status quo (utility pair $(0, 0)$) otherwise.

As discussed in the Introduction, it follows from Kalai and Samet's (1985) insight that applying a proportional rule to pick collective decisions guarantees the existence of a fully revealing BNE (involving even dominant strategies). Proportional rules were excluded from our analysis because they are not renegotiation-proof, a violation of our first regularity condition (Ex-Post Efficiency), and thus hard to enforce.

Even so, suppose that the group has a credible way to commit. In this section, we show that such rules rarely outperform our optimal regular rule in terms of the overall efficiency level. In other words, the inefficiency due to individuals withholding their types with our ex-post efficient rule is often lower than the inefficiency due to the fact that the symmetric proportional rule implements ex-post inefficient outcomes in some circumstances.

From the above description of the symmetric proportional rule, it follows that the efficiency loss is systematically equal to one in our benchmark model, and occurs if and only if $x \leq 1/2$ and $y \geq 1/2$, or $x \geq 1/2$ and $y \leq 1/2$. The efficiency loss is thus equal to

$$2F(1/2)(1 - F(1/2)). \tag{1}$$

If a regular rule is used, then inefficiency occurs when and only when neither individual discloses the collective action he is aware of. The total surplus being normalized to one in our benchmark model, the efficiency loss is thus equal to the probability of both individuals withholding their information at equilibrium, which is equal to the square of the equilibrium threshold (see Proposition 1 in the main paper). The inefficiency level is minimized when using the Nash solution (see Corollary 1 in the main paper). The associated efficiency loss is equal to $F(\theta_N)^2$.

We established in the main paper that $\theta_N \leq 1/2$, for all distribution f. Hence the efficiency loss associated to the Nash solution is at most $F(1/2)^2$. This is smaller than the efficiency loss associated to the symmetric proportional rule, see expression (1), whenever $F(1/2) \leq 2/3$. So, our optimal regular rule outperfoms the symmetric proportional solution whenever the likelihood of knowing a unfavorable collective action is at most 2/3.[4] Clearly θ_N often falls significantly below 1/2, and so the probability of inefficiency associated to the optimal regular solution is often much lower than $F(1/2)$, and the Nash solution also outperforms the symmetric proportional solution for many distributions f such that $F(1/2) > 2/3$.

[4]Put differently, the optimal regular rule outperforms the proportional rule if the likelihood that a player observes an unfavorable option is not substantially higher than his likelihood of observing a favorable option.