



**THE PINHAS SAPIR CENTER FOR DEVELOPMENT  
TEL AVIV UNIVERSITY**

"Strategic Disclosure of Collective Actions"

Geoffrey de Clippel<sup>1</sup>, Kfir Eliaz<sup>2</sup>

Discussion Paper No. 11-13

December 2013

**The paper can be downloaded from: <http://sapir.tau.ac.il>**

Thanks to The Pinhas Sapir Center for Development, Tel Aviv University for their financial support.

Financial support from the Deutsche Bank through the Institute for Advanced Study is gratefully acknowledged.

---

<sup>1</sup> Geoffrey de Clippel – Department of Economics, Brown University. Email: declippel@brown.edu

<sup>2</sup> Kfir Eliaz – The Eitan Berglas School of Economics, Tel Aviv University. Email: kfire@post.tau.ac.il

## Abstract

This paper illustrates how decision makers may have an incentive to withhold information regarding the existence of Pareto improving collective actions. The resulting level of inefficiency varies with the way compromises are reached when the parties have to choose among multiple options. Various reasonable compromise rules can be ranked unequivocally, and a rule resulting in a minimal level of inefficiency is identified. Qualitative results extend to sequential disclosure. Enforcing a hard deadline for disclosure may be welfare improving in some circumstances.

# 1 Introduction

Suppose some individuals (e.g. members of an organization, or bargainers) hold private and verifiable information regarding the availability of various collective actions. What are their incentives to disclose such information? We argue that withholding one's information may sometimes be preferable even if it is Pareto improving. Indeed, one must take into account the possibility that another participant may disclose an alternative action that is feasible and relatively more advantageous to oneself. We study how incentives to disclose vary with the way compromises are reached when parties must choose among multiple collective actions.

Such considerations are relevant, for instance, when a committee decides on hiring a candidate while it is not commonly known which individuals are available and/or suitable for the task (e.g. members of a research department hiring a junior colleague). Another example is that of a husband and a wife, who need to decide which house to buy, or which city/state/country to move to, or which school to send their kids to. Both sides need to identify feasible alternatives (e.g., which houses are for sale, which locations have relevant job openings, which schools have open slots), and the two may have conflicting interests regarding the choice (e.g., each may want a house closer to his/her workplace, each may want to move to a location with better career opportunities or closer to one's family, each may have a different preference on public versus private education). The above considerations would also be relevant in the case of countries looking for ways to address the development of a nuclear program by a hostile country, or to fight against terrorism, or to resolve a conflict.

In the private sector, popular techniques of “mutual gains” bargaining<sup>1</sup> include a step where both parties are asked to suggest possible solutions, before implementing an agreed-upon objective criterion to evaluate them. Circumstantial evidence suggests that strategic withholding is viewed as a possible

---

<sup>1</sup>Also known as interest-based, integrative, win-win, or principled bargaining, see e.g. Fisher et al. (1991), or the webpage of the Federal Mediation and Conciliation Service.

concern at this step. One vivid example appears in Haynes (1986), who discusses the role of mediators when implementing a mutual-gains approach to divorce and family issues:

*If the mediator determines that the parties are withholding options with a covert strategy in mind, the mediator can cite a similar situation with another couple and describe different options they considered. This can help break the logjam by forcing the couple to examine the options and including them on their list, thereby creating a greater level of safety for other options that are developed after one goes up on the board.<sup>2</sup>*

It is perhaps to limit strategic considerations that disputing parties are instructed to suggest possible solutions as rapidly as possible. It is also often emphasized that participants should refrain from evaluating options before a satisfactory number of suggestions have been made or parties have exhausted their ideas.

We start by studying the simplest model capturing the relevant strategic considerations we aim to illustrate. Two individuals are each privately aware of the feasibility of a collective action resulting in utility pairs that fall on the straight line joining  $(1, 0)$  to  $(0, 1)$ . Later (see Section 6), we indicate how our results extend to more general frameworks, including situations where individuals may be unaware of any feasible action with positive probability, situations where individuals may be aware of more than one feasible action, and situations where collective actions result in utility pairs that fall on a symmetric curve that need not be straight.<sup>3</sup>

In the static disclosure game, participants must decide simultaneously whether to disclose the collective action they are aware of. Depending on the strategy they follow, and the hard evidence they hold, either zero, one

---

<sup>2</sup>In our model, though, feasible options can only be disclosed by the bargainers themselves - there will be no mediator with extra information to break the logjam.

<sup>3</sup>Extensions to asymmetric problems, problems involving more than two individuals, and problems with feasible actions that result in Pareto comparable utility profiles are left for future work.

or two collective actions will be disclosed. The final outcome and resulting utilities then depend on how collective decisions are made. Various procedures come to mind. Perhaps the simplest is to choose at random with the help of a fair coin. More generally, collective decisions could result from the implementation of given bargaining protocols (see e.g. Rubinstein’s (1982) alternating offers), or the application of basic fairness principles. Our analysis accommodates diverse ways of reaching compromises. Formally, a compromise rule determines final utility pairs as a function of the actions that have been disclosed. It is “regular” if it satisfies basic properties of anonymity, ex-post efficiency, and monotonicity (see definitions in Section 2). Proposition 1 provides a characterization of the (unique) symmetric Bayesian Nash equilibrium of the disclosure game given any regular compromise rule.

Our second main result is normative in nature. Are some regular rules unequivocally superior to others in terms of the efficiency level they induce in the disclosure game? We define a partial order on compromise rules, and prove that the efficiency level is increasing with that order. It also allows to identify an optimal regular rule, which happens to coincide with the Nash (1950) solution in the special case where utility pairs associated with collective decisions fall on the straight line joining on  $(1, 0)$  and  $(0, 1)$ . In addition, the partial ordering induces a lattice structure on the set of regular compromise rules: given any pair of regular rules, we can construct a new pair of regular rules, one which dominates the original rules, and another which is inferior.

It is tempting to think that inefficiency is driven by the fact that there is a single opportunity to speak. Wouldn’t individuals give in, and disclose their information, right before they are about to settle on the Pareto inferior status quo because no collective action has been disclosed? Things are not that simple. At any point in time, an individual  $i$  would prefer  $j$  to give in first when there is a high likelihood that  $j$  is aware of a collective action that is relatively more advantageous to  $i$ . One must consider a dynamic disclosure game to address this question. Both individuals now choose when to disclose their information, if at all. A collective decision is made as soon as at least one action has been disclosed, using the compromise rule whenever multiple actions

are disclosed at the same time. Our third main result offers a characterization of the perfect Bayesian Nash equilibrium for this game. While all types get disclosed at some point in time, inefficiency remains in the form of delay. The time at which types are disclosed depend on the compromise rule that is being implemented. Our fourth main result shows that the partial ranking we identified in our analysis of the static game remains relevant, as information gets disclosed no later when using a superior compromise rule. In particular, the compromise rule that was optimal for the static game remains optimal for the dynamic game as well.

Some types that were not disclosed in the static game are now disclosed with delay in the dynamic game, which has a positive impact on welfare. On the other hand, some other types that were disclosed in the static game are now disclosed with delay in the dynamic game, which has a negative impact on welfare. Welfare comparisons between the static and the dynamic procedures are thus not immediate. We provide a robust example where enforcing a hard deadline, namely to speak once or never thereafter, is preferable to letting individuals speak at any point in time.

We also analyze a variant of our dynamic disclosure game, where individuals have an opportunity to react by disclosing the collective action they are aware of right after the other participant disclosed his, in which case the compromise rule is applied. Our fifth main result shows that any regular compromise rule induces the same perfect Bayesian Nash equilibrium, and that this equilibrium is inferior to the equilibrium that prevails when individuals do not have the opportunity to react. This provides yet another illustration of the fact that limiting the opportunity to speak may have a positive impact on welfare.

## **Related Literature**

Our paper fits into the literature on information revelation in committees (see e.g. Hao and Suen (2009) for a survey), as well as the literature on expert advices (see e.g. Sobel (2010) for a survey) if one interprets the compromise rule as capturing the decision of a fictitious third party. Milgrom and Roberts

(1986) is the closest model within this literature as it shares two important features with our setting, the presence of multiple “senders” (the two informed individuals) and of verifiable information (a problem of disclosure instead of cheap talk). Some key differences, though, lead to distinct results. Competition among multiple senders leads to first-best efficiency in their model, but not in ours because information is incomplete. An individual does not know indeed the collective action that his partner is aware of. Classic results on information unraveling do not apply either because information is about the feasibility of collective actions, not about a state of nature that impacts preferences. If an individual knows that his opponent is withholding information regarding the feasibility of some collective action, there is nothing he or the third party can do without knowing what that collective action is.

In a classic paper, Kalai and Samet (1985) study a noncooperative destruction game that precedes a bargaining stage. Opportunities at the bargaining stage are captured in their model by utility possibility sets. In the prebargaining stage, individuals simultaneously choose which utility pairs (if any) to take away from that set. A proportional solution for the bargaining stage associates to each utility possibility set the utility pair in it that is as high as possible on a given straight line with positive slope. Kalai and Samet observe that no destruction is a dominant strategy for both bargainers when bargaining outcomes coincide with a fixed proportional solution.<sup>4</sup> While our game involves disclosure instead of destruction, and is of incomplete instead of complete information, a similar result applies in our framework: compromising via a given proportional solution makes disclosure dominant in all circumstances. Proportional compromise rules are not regular (and thus not covered by our analysis) because they are ex-post inefficient. Rules that are ex-post inefficient are unlikely to be followed, as they are not renegotiation-proof. For instance, the proportional rule selects the status quo when only one individual reveals his

---

<sup>4</sup>Kalai and Samet also establish a partial converse result when utility possibility sets are comprehensive: (a) no destruction is a dominant strategy in all problems only if the solution is monotone (a different notion of monotonicity than the one we consider below), and (b) a solution is monotone, (weakly) Pareto efficient, homogenous and strictly individually rational if and only if it is proportional (see also Kalai (1977)).

type in the disclosure game,<sup>5</sup> despite the fact that a Pareto superior collective action is available. A similar problem occurs when both individuals disclose collective actions that deliver utility pairs that are both above or both below the straight line defining the proportional solution. Beyond the difficulty associated with a lack of renegotiation-proofness, we show in the Supplementary Appendix that the efficiency loss associated with proportional rules (due to inefficiency at the compromise stage) is larger than the efficiency loss of our optimal ex-post efficient rule (due to inefficiency at the disclosure stage), except in rare circumstances where individuals face a high likelihood of knowing a collective action that is “unfavorable” (see Definition 1) to them.

Frankel (1998) provides a first analysis of individual incentives to identify collective actions, a problem of acquisition instead of transmission of information. For instance, in one of the models he studies, each bargainer  $i$  simultaneously chooses a probability  $p_i$  and an interval  $I_i \subseteq [\frac{1}{2}, 1]$ , incurring a cost, which increases with  $p_i$  and decreases with the length of  $I_i$ . With probability  $p_i$  bargainer  $i$  gets an “idea”, which gives him a payoff of  $x_i$ , where  $x_i$  is drawn from a uniform distribution on  $I_i$ . With probability  $1 - p_i$ , bargainer  $i$  gets no idea. The realizations of the two bargainers are then revealed (they have no choice on the matter) and a coin is tossed to select one of the realized ideas (if only one idea was realized, it is selected with certainty), and if no idea was realized, both get zero. Frankel shows that in equilibrium, the choices of  $(p_i, I_i)_{i=1,2}$  can be either excessive or suboptimal. Determining how different compromise rules affect individual incentives to acquire costly information regarding the feasibility of collective actions, and building on our model by adding costly search as a preliminary stage to strategic disclosure are both interesting questions for future research.

The dynamic disclosure game, presented in Section 5, is related to the analysis of the war of attrition and related games of costly waiting in continuous time and with asymmetric information (see e.g., Gul and Lundholm (1995), Ponsati and Sákovics (1995) and Bulow and Klemperer (1999)). The key dif-

---

<sup>5</sup>Except in the zero probability event where the disclosed collective action results in utilities with the right proportions



ference between these games and ours is that each player’s payoff depends on the other player’s type, either directly when one’s opponent has disclosed a collective action, or via the compromise rule when both have disclosed at the same time (as may happen in the unique symmetric equilibrium of our game). In addition, the war of attrition has oftentimes been analyzed as an all-pay auction where the players’ preferences are quasi-linear in the cost of delay (see, e.g., the “generalized” war of attrition by Bulow and Klemperer (1999)), which is not the case in our framework. These differences imply that standard techniques used to solve the above waiting games do not apply to the present context.

## 2 Benchmark Model

Two individuals face a large set of collective actions, but do not know which are actually feasible and/or what the associated payoffs are. Each has learned, either by chance or as a result of active search,<sup>6</sup> about the feasibility of a single collective action, represented in the space of utilities as a pair of non-negative real numbers. As a starting point, we assume that these payoffs fall on the straight line joining  $(1, 0)$  to  $(0, 1)$ , with  $(0, 0)$  determining the reference payoffs that prevail when no collective action is implemented. We show in Section 6 that our analysis extends to more general sets, and to problems where an individual may be aware of multiple actions or none at all.

To simplify the exposition we define an individual’s type to be his utility from the payoff pair he is aware of. We let  $x$  denote the first individual’s type (i.e., he is aware of the payoff pair  $(x, 1 - x)$ ), and we let  $y$  denote the other individual’s type (i.e., he is aware of  $(1 - y, y)$ ).

An individual does not know what utility pair his opponent is aware of. Beliefs are captured by a common density  $f$  on  $[0, 1]$  with full support. Consider for instance the case of two parties with conflicting objectives who need to agree on a person to hire. While parties may know the distribution of the potential candidates’ characteristics in the population, the characteristics of

---

<sup>6</sup>In this latter case, our work can serve as a building block of a more elaborate model designed to understand incentives to search before the disclosure stage.

any given candidate and whether he or she is interested in being considered for the position are not self-evident.

The game starts with the *disclosure stage*. At that point, individuals decide independently whether to disclose the feasibility of the collective action they are aware of.<sup>7</sup> We assume that types are verifiable once disclosed, and hence an individual cannot report anything else than what he knows. The benchmark model may be viewed as a situation in which the two parties need to submit hard evidence on a feasible option (from which the payoffs can be inferred) until some prespecified (and rigid) deadline, and submissions will be reviewed only at the deadline (i.e., submitting before the deadline has no effect, and evidence submitted after the deadline is not accepted).

The game ends with the *compromise stage*, where individuals decide which collective action to implement. Only those actions for which there is verifiable evidence attesting to their feasibility (and from which the payoffs can be inferred) can be selected. Going back to our hiring example, members of the hiring committee can only choose among a list of candidates that is presented to them. Similarly, when heads of countries meet to decide on a response to terrorism, they will only consider those concrete plans of actions that were presented to them.

Including the possibility of keeping the status quo, there are one, two or three options to choose from depending on the number of actions that have been disclosed. We study problems where monetary compensations are not available, or deemed inappropriate. Compromises can thus be reached only by using lotteries. The details of the underlying process determining how this lottery depends on the set of available options does not matter for our analysis. All what matters is how the final payoffs vary, which will be captured by a *compromise rule*  $r : ([0, 1] \times \{\emptyset\}) \times ([0, 1] \times \{\emptyset\}) \rightarrow \mathbb{R}_+^2$ . The compromise rule can be a reduced-form of the equilibrium outcome of some explicit bargaining protocol, or of the outcome associated with some fairness criteria, or of some previously agreed-upon standard.

---

<sup>7</sup>We assume that no exogenously imposed sequence of moves can be enforced. In Section 5 we endogenize the sequence of moves by letting individuals decide if and when to disclose.

Our analysis applies to any compromise rule  $r$  that satisfies the following three regularity properties. A *regular* compromise rule is

1. *ex-post efficient*, meaning that it selects utility pairs that are Pareto efficient given the collective actions that have been disclosed: for all  $x, y \in [0, 1]$ ,  $r(x, \emptyset) = (x, 1 - x)$ ,  $r(\emptyset, y) = (1 - y, y)$ , and there exists  $\alpha \in [0, 1]$  such that  $r(x, y) = \alpha(x, 1 - x) + (1 - \alpha)(1 - y, y)$ .
2. *anonymous*, meaning that the individuals' identity is irrelevant for determining the compromise:  $r_i(x, y) = r_j(y, x)$ .
3. *monotone*, in the sense of giving a larger utility to an individual when disclosed actions are more favorable to him:  $x' \geq x, y' \leq y \Rightarrow r_1(x', y') \geq r_1(x, y)$  and  $r_2(x', y') \leq r_2(x, y)$ , for all  $x, x', y$ , and  $y'$  in  $[0, 1]$ .

The regularity conditions are meant to capture common features of prevalent collective decision processes. The first part of our analysis will be positive, characterizing symmetric Bayesian Nash equilibria of the disclosure game for a given regular compromise rule, and computing their associated level of efficiency. For this part, we highlight that our regularity conditions are met by all classic collective decision processes discussed in the literature (non-cooperative and axiomatic bargaining, as well as social choice). The second part of our analysis is normative, characterizing a regular compromise rule that systematically minimizes the level of inefficiency at the disclosure stage. This could be viewed as a constrained mechanism design exercise. Even if they might be ready to change to some extent the way compromises are reached in order to improve efficiency at the disclosure stage, they are not ready to give up on these basic properties that are of primary importance to them.

**Example 1** *Ex-post efficiency fully determines which collective action to implement when only one has been disclosed. Compromise rules thus vary only in how the final outcome is determined when two collective decisions have been disclosed. The simplest way to reach a compromise in such circumstances is to decide which action to implement with the help of a fair coin.<sup>8</sup> The coin-flip*

---

<sup>8</sup>This rule coincides to Raiffa's (1957, Section 6.7) solution in bargaining theory.

rule is thus defined by

$$r_{CF}(x, y) = \left( \frac{x + (1 - y)}{2}, \frac{(1 - x) + y}{2} \right)$$

for all  $x, y \in [0, 1]$ .

A second classic way to resolve conflict is captured by Nash's (1950) bargaining solution. It is obtained by maximizing the product of the two individuals' utilities. In our benchmark model, this rule picks the lottery that brings the individuals' utilities as close as possible to  $(1/2, 1/2)$ . Formally:

$$r_N(x, y) = \begin{cases} (\max\{x, 1 - y\}, 1 - \max\{x, 1 - y\}) & \text{if } \max\{x, 1 - y\} \leq \frac{1}{2} \\ (\min\{x, 1 - y\}, 1 - \min\{x, 1 - y\}) & \text{if } \min\{x, 1 - y\} \geq \frac{1}{2} \\ (\frac{1}{2}, \frac{1}{2}) & \text{otherwise} \end{cases}$$

A third bargaining solution that is often discussed in the literature was proposed by Kalai and Smorodinsky (1975). This rule also appears in social choice under the name of relative egalitarianism. Indeed, it picks the Pareto efficient utility profile that equalizes the individuals' utility gains, relative to the maximal utility they could achieve. Formally:

$$r_{KS}(x, y) = \left( \frac{\max(x, 1 - y)}{\max(x, 1 - y) + \max(1 - x, y)}, \frac{\max(1 - x, y)}{\max(x, 1 - y) + \max(1 - x, y)} \right).$$

### 3 Positive Analysis of the Disclosure Stage

A mixed-strategy for individual  $i$  in the disclosure stage is a measurable function  $\sigma_i : [0, 1] \rightarrow [0, 1]$ , where  $\sigma_i(x)$  is the probability that type  $x$  of individual  $i$  announces the collective action he is aware of. A pair of mixed strategies, one for each individual, forms a *Bayesian Nash equilibrium* (BNE) of the disclosure game if the action it prescribes to each type of each individual is optimal against the opponent's strategy. The BNE is *symmetric* if both individuals follow the same strategy.

**Proposition 1** *The disclosure game admits a unique symmetric BNE, in which each individual discloses his type if and only if it is greater or equal*

to a strictly positive threshold

$$\theta = \sup\{x \in [0, \frac{1}{2}] \mid xF(x) + \int_{y=x}^1 (r_1(x, y) - (1 - y))f(y)dy < 0\}.$$
 <sup>9</sup> (1)

The associated equilibrium outcome is thus inefficient.

The proof relies on properties of the function determining the expected net gain of revealing over withholding one's type given the opponent's strategy. For individual 1 this is given by

$$ENG_1(x, \sigma_2) = x \int_{y=0}^1 (1 - \sigma_2(y))f(y)dy + \int_{y=0}^1 \sigma_2(y)[r_1(x, y) - (1 - y)]f(y)dy,$$
 (2)

for each type  $x \in [0, 1]$  and each strategy  $\sigma_2$  (ENG for is defined analogously for the second individual). It is not difficult to check (see Lemma 1 in the Appendix) that ENG increases with types (independently of the other individual's strategy), and strictly decreases as the other individual becomes more likely to disclose each of his types. As a consequence, BNEs must involve threshold strategies, and equilibrium existence is proved using Brouwer's fixed point theorem.<sup>10</sup>

If the symmetric BNE is the unique BNE (as is the case, for instance, when applying the coin-flip rule with a uniform distribution of types), then it is the only rationalizable strategy profile.

**Proposition 2** *If the symmetric BNE is the unique BNE, then it is also the unique profile of strategies that survives the iterated elimination of strictly dominated strategies.*

Our analysis focuses on symmetric BNEs. This is natural given that individuals are symmetric (ex-ante) with  $r$  being anonymous, and types being drawn from a same distribution. The disclosure game may also have asymmetric BNEs in addition to the unique symmetric one. Inefficiency often prevails at all those equilibria as well. A compromise rule is *symmetric* if the pairs of utilities it selects depends only on the payoffs associated with the collective

---

<sup>9</sup>If  $r$  is continuous, then the threshold  $\theta$  is given by the solution to the following simpler equation:  $\theta F(\theta) + \int_{y=\theta}^1 (r_1(\theta, y) - (1 - y))f(y)dy = 0$ .

<sup>10</sup>Observe that existence is obtained without requiring compromise rules to be continuous.

actions that have been disclosed, and not on the identity of who has disclosed what:  $r(x, y) = r(1 - y, 1 - x)$ . This property is satisfied by all classic compromise rules, and in particular by all the compromise rules provided as examples in this paper.

**Proposition 3** *Let  $r$  be a regular compromise rule that is continuous and symmetric. If it is more likely to discover collective decisions that are relatively more favorable, in the sense that  $f(x) > f(1 - x)$ , whenever  $x \geq 1/2$ , then the outcome associated with any BNE of the disclosure game associated with  $r$  is inefficient.*

## 4 How to Promote Disclosure?

Inefficiency arises in our setting when no individual reveals his type in the disclosure stage. The expected level of inefficiency is thus equal to the probability of that event, which is the square of the threshold characterized in equation (1). The Supplementary Appendix contains computations of this threshold for  $r_{CF}$ ,  $r_{KS}$ , and  $r_N$  in the case of a uniform distribution of types. The resulting numbers are 0.236, 0.22, and 0.183 respectively. Hence we see that using the Nash solution as a compromise rule is welfare improving compared to Kalai-Smorodinsky, which in turn is preferable to the coin-flip rule. Can we rank other compromise rules? Are there alternatives that are preferable to  $r_N$ ? Do results depend on the specific distribution of types?

In order to address these questions, we introduce a partial ordering on compromise rules.

**Definition 1** *An individual's type is unfavorable if the payoff that is relevant for him is lower than  $1/2$ . The first individual is in a weaker position if  $x \leq y$ . Similarly, the second individual is in a weaker position if  $y \leq x$ .*

Notice that  $x \leq y$  is equivalent to  $\min\{x, 1 - y\} \leq \min\{1 - x, y\}$ . In other words, the first individual is in a weaker position if the worst payoff he can get, given the disclosed collective actions, is lower than the worst payoff of the other individual.<sup>11</sup>

<sup>11</sup>Note that all regular compromise rules are consistent with the concept of weak position,

**Definition 2** *The compromise rule  $r'$  dominates the compromise rule  $r$  if  $r'$  gives a higher payoff than  $r$  to an individual with an unfavorable type who is in a weaker position. Formally,  $r' \succeq r$  if  $r'_1(x, y) \geq r_1(x, y)$ , for all  $x, y$  such that  $x \leq \min\{1/2, y\}$ , and  $r'_2(x, y) \geq r_2(x, y)$ , for all  $x, y$  such that  $y \leq \min\{1/2, x\}$ .*

**Proposition 4** *If the compromise rule  $r'$  dominates the compromise rule  $r$ , then the probability of inefficiency in the symmetric equilibrium of the disclosure game associated with  $r'$  is smaller or equal to the probability of inefficiency in the symmetric equilibrium of the disclosure game associated with  $r$ .*

Proposition 4 has multiple implications. First it allows to identify a regular rule that is optimal in terms of induced efficiency in the disclosure game.

**Corollary 1**  *$r_N \succeq r$ , for any regular compromise rule  $r$ . Hence the probability of inefficiency in the symmetric equilibrium of the disclosure game associated with any regular compromise rule is larger or equal to the probability of inefficiency in the symmetric equilibrium of the disclosure game associated with the Nash solution.*

One may get the impression that the Nash solution is optimal simply because it is most favorable to individuals in a weaker position among all regular rules. This argument is too simplistic to constitute a proof, as favoring weaker types also reduces incentives to disclose conditional on the opponent disclosing a collective action that is even more unfavorable to him. So it needs not be the case that favoring weaker types guarantees a higher probability of disclosure for each type. Proving optimality requires understanding equilibrium behavior. To illustrate this point, consider the *deterministic Nash* solution that selects the utility pair that is most favorable to the individual in the weaker position. By definition, this compromise rule favors the individual in the weak position, even more so than  $\sigma_N$ , but at the cost of violating the third regularity condition (monotonicity).<sup>12</sup> When  $f$  is uniform, the one-shot disclosure game

---

giving a smaller payoff to the individual who is in a weaker position. Indeed, suppose, for instance, that 1 is weaker, i.e.  $x \leq y$ . Anonymity implies that  $r_1(x, x) = 1/2$ . Monotonicity implies that  $r_1(x, y) \leq 1/2$ , and hence  $r_1(x, y) \leq r_2(x, y)$ .

<sup>12</sup>Indeed, it picks  $(1/2, 1/2)$  out of  $\{(0, 0), (1/3, 2/3), (1/2, 1/2)\}$ , and  $(1/3, 2/3)$  out of  $\{(0, 0), (1/3, 2/3), (3/4, 1/4)\}$ . The first individual's payoff decreases, while disclosed utility pairs become more favorable to him.

induced by the deterministic Nash solution has a symmetric BNE where the probability of disclosure in equilibrium is given by

$$\sigma(x) = \begin{cases} -\frac{1}{3} + \frac{4}{3}\left(\frac{1}{2-4x}\right)^{\frac{3}{2}} & \text{if } x \leq \frac{1}{4} \\ 1 & \text{if } x > \frac{1}{4} \end{cases}$$

Applying  $\sigma_N$ , on the other hand, results in individuals disclosing types above 0.183. We see that favoring more the individual in the weak position does not make each type more likely to disclose at equilibrium.

A dual to Corollary 1 gives an upper bound on the probability of inefficiency associated with any regular compromise rule. Consider the compromise rule that maximizes the maximum of the two individuals' payoffs,

$$r_{MM}(x, y) = \begin{cases} (x, 1-x) & \text{if } \max\{x, 1-x\} > \max\{1-y, y\} \\ (1-y, y) & \text{if } \max\{x, 1-x\} < \max\{1-y, y\} \\ (1/2, 1/2) & \text{if } x = y \end{cases}$$

(in other words, this solution picks the point that is the furthest from  $(1/2, 1/2)$ , i.e., it minimizes the product of the individuals' payoffs). It is easy to check that  $r_{MM}$  is regular, and that  $r \succeq r_{MM}$ , for any regular compromise rule  $r$ . By Proposition 4, the probability of inefficiency in the symmetric equilibrium of the disclosure game induced by any regular compromise rule is smaller or equal to the probability of inefficiency in the symmetric equilibrium of the game induced by  $r_{MM}$ .<sup>13</sup>

We show in the Supplementary Appendix that  $r_{KS} \succeq r_{CF}$ . Proposition 4 implies that the equilibrium outcome associated with the Kalai-Smorodinsky solution is never less efficient than the one associated with the coin-flip rule.

Proposition 4 also suggests an algorithm that transforms any pair of regular rules into a regular rule, whose associated symmetric BNE is at least as efficient as the symmetric BNE of each of the two original rules. A similar procedure yields a regular rule, which results in a lower level of efficiency. In other words, the partial ordering of solutions according to their efficiency induces a lattice structure over regular compromise rules.

---

<sup>13</sup>For instance, the equilibrium strategies' threshold is equal to 0.293 in the case of a uniform distribution on types.



For any pair of regular compromise rules,  $r$  and  $r'$ , let  $r \vee r'$  be the compromise rule defined as follows. First,

$$(r \vee r')_1(x, y) = \begin{cases} \max\{r_1(x, y), r'_1(x, y)\} & \text{if } x \leq y \\ \min\{r_1(x, y), r'_1(x, y)\} & \text{if } x \geq y \end{cases}$$

where  $(r \vee r')_2(x, y) = 1 - (r \vee r')_1(x, y)$ . Second,  $(r \vee r')(x, \emptyset) = (x, 1 - x)$  and similarly,  $(r \vee r')(\emptyset, y) = (1 - y, y)$ . In an analogous way we define  $r \wedge r'$ , where the only difference is the following: if the first individual is in a weaker position, then  $(r \wedge r')_1(x, y)$  equals  $\min\{r_1(x, y), r'_1(x, y)\}$ , and if the second individual is in a weaker position, then  $(r \wedge r')_1(x, y)$  equals  $\max\{r_1(x, y), r'_1(x, y)\}$ .

**Proposition 5** (i)  $r \vee r'$  and  $r \wedge r'$  are regular rules, (ii)  $r \vee r' \succeq r$  and  $r \vee r' \succeq r'$ , and (iii)  $r \succeq r \wedge r'$  and  $r' \succeq r \wedge r'$ .

We conclude this section by discussing asymmetric BNEs. The distribution  $f$  is symmetric if types  $x$  and  $1 - x$  are equally likely:  $f(x) = f(1 - x)$ , for all  $x \in [0, 1]$ . The Nash solution is optimal for such distributions, not only when considering the more natural symmetric BNEs, but also when considering asymmetric BNEs.

**Proposition 6** Let  $r$  be a regular compromise rule that is continuous and symmetric. If  $f$  is symmetric as well, then efficiency in the disclosure game associated with  $r$  is obtained at some BNE only if  $r$  is the Nash solution.

## 5 Dynamic Disclosure Game

In the dynamic disclosure game, individuals decide when to disclose their type, if at all. The compromise rule is applied as soon as at least one collective action has been disclosed (we also study below the case where individuals have an opportunity to react before the compromise rule is applied). We restrict attention to symmetric pure-strategy Bayesian Nash equilibria. A strategy is a measurable function  $\tau : [0, 1] \rightarrow \mathbb{R}_+ \cup \{\infty\}$ , which determines for each type  $x$  the time  $\tau(x)$  at which to reveal  $x$ .<sup>14</sup> Measurability means that the inverse image of any Lebesgue measurable set (in particular any interval) is Lebesgue

---

<sup>14</sup> $\tau(x) = \infty$  means that the individual never discloses when of type  $x$ .

measurable:  $\tau^{-1}(T) = \{x \in [0, 1] | \tau(x) \in T\}$  is Lebesgue measurable if  $T$  is Lebesgue measurable. It guarantees that an individual's expected utility when his opponent is known to reveal over some given interval of time, is well-defined. Utilities are discounted exponentially over time following a discount factor  $\delta < 1$ . The outcome when the first individual is of type  $x$ , while the other individual is of type  $y$ , and they both implement the strategy  $\tau$ , is  $x$  at time  $\tau(x)$  if  $\tau(x) < \tau(y)$ ,  $y$  at time  $\tau(y)$  if  $\tau(x) > \tau(y)$ , and  $r(x, y)$  at time  $\tau(x)$  if  $\tau(x) = \tau(y)$ .<sup>15</sup>

## 5.1 Positive Analysis

The strategy  $\tau$  is part of a *symmetric Bayesian Nash equilibrium* if, for every type  $x \in [0, 1]$ , the expected net gain of revealing at any time  $t \geq 0$  different from  $\tau(x)$  is non-positive, where an individual's expected net gain - let's say the first individual to fix ideas - is given by the following formula when  $t > \tau(x)$  (a similar formula applies in the other case):

$$\begin{aligned} ENG_1(t \text{ vs. } \tau(x), x) &= x(e^{-\delta t} - e^{-\delta\tau(x)}) \int_{y \in \tau^{-1}([t, \infty])} f(y) dy \\ &+ \int_{y \in \tau^{-1}(t)} (e^{-\delta t} r_1(x, y) - e^{-\delta\tau(x)} x) f(y) dy \\ &+ \int_{y \in \tau^{-1}([\tau(x), t])} (e^{-\delta\tau(y)} (1 - y) - e^{-\delta\tau(x)} x) f(y) dy \\ &+ \int_{y \in \tau^{-1}(\tau(x))} e^{-\delta\tau(x)} ((1 - y) - r_1(x, y)) f(y) dy. \end{aligned}$$

We will need the following additional assumption on  $r$  to establish the uniqueness of the symmetric BNE:

$$r_1(x, \frac{1}{2}) < x, \forall x > 1/2, \text{ and } r_1(x, \frac{1}{2}) > x, \forall x < 1/2. \tag{3}$$

<sup>15</sup>The dynamic disclosure game is similar to classic wars of attrition, in that both parties incur a cost of delay when neither gives in. Our discussion of the related literature in the Introduction indicates key differences which makes it impossible to apply results from that literature to our disclosure problem.

<sup>16</sup>A similar pair of conditions necessarily hold for the second individual as well, as a

The weak inequality holds for any regular rule. Requiring a strict inequality is a mild additional requirement which is satisfied by all classic rules (Kalai-Smorodinsky, Nash and coin flip), but not the max-max rule ( $r_{MM}$ ).

**Proposition 7** *Let  $\tau^*$  be the strategy defined as follows:*

$$\tau^*(x) = \begin{cases} 0 & \text{if } x \geq \theta \\ \int_x^\theta \frac{(1-2y)f(y)}{\delta y F(y)} dy & \text{if } x < \theta, \end{cases}$$

where

$$\theta = \sup\{x \in [0, 1/2] \mid \int_{y=x}^1 (r_1(x, y) - (1-y))f(y)dy < 0\}. \quad (4)$$

*The pair of strategies  $(\tau^*, \tau^*)$  forms a symmetric Bayesian Nash equilibrium of the dynamic disclosure game. If  $r$  satisfies condition (3), then it is the unique symmetric BNE of the game.*

To understand the intuition for this result, consider first the threshold for immediate disclosure, as given by (4). Roughly speaking, the type at the threshold is indifferent between disclosing immediately and disclosing a “little” later, say at  $\varepsilon$ . The potential gain from waiting is that weaker types may disclose more favorable collective actions. The potential loss is that the game may end with the opponent disclosing before him a type which leads to a less favorable outcome. Since  $\varepsilon$  is sufficiently small, there is no loss from delay (hence, the discount factor does not appear in (4)), and almost all types who disclose after time zero, also disclose after time  $\varepsilon$ . This latter point explains why the term  $x F(x)$  (the payoff from types weaker than  $x$ ), does not appear in the equation for the dynamic cutoff (although it does appear in equation (1) that determines the static cutoff). The cutoff type  $\theta$  cannot be higher than  $\frac{1}{2}$  because the types who disclose immediately would give  $\theta$  a lower payoff than if he would compromise. On the other hand,  $\theta$  cannot be too low because there would be an incentive to delay disclosure as the other party’s type is likely to be lower than  $1 - \theta$ .

Note that the compromise rule affects only the threshold of types who disclose immediately. It has no effect on the rate at which types lower than  $\theta$  

---

consequence of the second regularity condition (Anonymity).

delay their disclosure. This follows from the fact that the compromise rule affects the outcome whenever there is a mass of types who disclose at the same time. This occurs at time zero where more than half of the types disclose. However, relatively unfavorable types, who delay their disclosure, have no incentive to pool and disclose at the same time. Intuitively, the most unfavorable type within the set of types who pool would have an incentive to delay his disclosure by a “little bit”: all the types he pooled with before would now disclose before him, and since their type is lower than  $\frac{1}{2}$ , he would get a payoff higher than  $\frac{1}{2}$  (and hence, also higher than his type). Proving that there is no mass of types who pool on a positive date requires rather involved arguments, and details are provided in the appendix.

A type who discloses with delay chooses the latest time at which to disclose by balancing the marginal benefit from delay with the marginal cost. This is reflected in the equilibrium strategy of types below the cutoff  $\theta$ . To understand this strategy consider a type  $x < \theta$ , who pretends to be a “slightly” lower type,  $x - \varepsilon$ . For each type  $x - \varepsilon < y < x$  of the other individual, the net expected gain from this deviation is  $((1 - y) - x)f(y)$ . The potential loss from this deviation is the wasteful delay that would occur if the other party’s type is even lower than  $x - \varepsilon$ . At the limit when  $\varepsilon$  goes to zero, the ratio of the expected marginal gain to the expected marginal loss is  $\frac{(1-2x)f(x)}{\delta x F(x)}$ , which is the derivative of the equilibrium strategy of type  $x$ .

To prove that  $\tau^*(x)$  is an equilibrium, we show that no type has an incentive to deviate, given that the other individual uses this strategy. The proof of uniqueness is more involved and details are provided in the appendix.

## 5.2 Normative Analysis

As evident from Proposition 7, the level of efficiency induced by a compromise rule depends entirely on the threshold. An important implication of this is that the partial ordering identified in Proposition 4 for the static game continues to predict the level of efficiency of equilibrium outcomes in the dynamic game as well.

**Proposition 8** *Let  $r$  and  $r'$  be two regular compromise rules that satisfy (3), and let  $\tau$  and  $\tau'$  be the strategies in the symmetric BNE of the dynamic disclosure game associated with  $r$  and  $r'$  respectively. If  $r' \succeq r$ , then  $\tau'(x) \leq \tau(x)$ , for each  $x \in [0, 1]$ , and the equilibrium outcome associated with  $r'$  is Pareto superior.*

Corollary 1 and Proposition 5 thus extend to the dynamic game.

### 5.3 Dynamic vs. Static

We start our comparison of the dynamic and static versions of the disclosure game by showing that, for any regular compromise rule satisfying condition (3), some types that are being disclosed in the static game are not getting disclosed right away in the dynamic game. In other words, the threshold derived for the static game is lower than the threshold derived for the dynamic game. Let  $\theta_S$  and  $\theta_D$  be the thresholds given by (1) and (4), respectively (i.e., the former is the cutoff of the one-shot simultaneous game, while the latter is the cutoff of the dynamic game).

**Proposition 9** *If  $r$  is a regular compromise rule satisfying (3), then  $\theta_S \leq \theta_D$ .*

Proposition 9 raises the following question: given a regular compromise rule satisfying (3), are individuals better off in the symmetric BNE of the static game or the dynamic game? We address this question in the special case where  $f$  is uniform.<sup>17</sup>

**Proposition 10** *Suppose that the distribution of types is uniform. Let  $r$  be a regular compromise rule that satisfies (3). Each individual's ex-ante expected payoff is equal to  $\frac{1-\theta_S^2}{2}$  in the static game, and equal to*

$$\frac{1}{2}[1 - \theta_D^2 - 2e^{1/\theta_D} \cdot \theta_D^2 \cdot E_i(-\frac{1}{\theta_D})] \quad (5)$$

*in the dynamic game, where  $E_i(x)$  denotes the exponential integral. As a consequence, the outcome of the dynamic game Pareto dominates that of the static game for  $r = r_N, r_{KS}$ , and  $r_{CF}$ .*

<sup>17</sup>It remains an open question whether the superiority of the static disclosure game remains true for all regular compromise rules and for all distributions.

## 5.4 Dynamic Disclosure with an Opportunity to React

As a natural variant of our dynamic game, we study a situation where individuals have one last chance to disclose the collective action they are aware of right after the other has “spoken”, i.e. right before  $r$  is implemented. Note that the strategies in this game are richer than those of the original dynamic game. As in the original game, they specify the latest period in which an individual would disclose if the other party has not done so. But in addition, for every history which ended with disclosure by the other party, an individual’s strategy also specifies whether or not he would disclose as a function of the other party’s disclosed type and the period of disclosure. To eliminate notational complications and unlikely off-equilibrium behavior, we focus on a slightly refined notion of BNE. Indeed, we will assume that type  $x$  discloses right after the other party has disclosed a type  $y$  if and only if  $y > 1 - x$ . In other words, we focus on equilibria in which an individual discloses immediately after the other party has disclosed whenever it is optimal for him to do so (whenever the the payoff from the other party’s collective action is lower than the payoff from his own). Given this restriction, strategies in a refined BNE are measurable functions  $\tau : [0, 1] \rightarrow \mathbb{R}_+ \cup \{\infty\}$ , that describe when an individual discloses as a function of his type.

**Proposition 11** *The modified dynamic disclosure game admits a unique refined symmetric Bayesian Nash equilibrium. The equilibrium disclosure strategy  $\mathbf{t}^*$  for both individuals is the following:*

$$\mathbf{t}^*(x) = \begin{cases} 0 & \text{if } x \geq 1/2 \\ \int_x^{1/2} \frac{(1-2y)f(y)}{\delta y F(y)} dy & \text{if } x \leq 1/2. \end{cases}$$

By Proposition 11, the timing of disclosure in the unique refined symmetric BNE is *independent* of the compromise rule. As in the original dynamic game, types who delay have no incentive to pool. Hence, the rate of delay is independent of the compromise rule. The reason the threshold type is independent of the compromise rule is that the opportunity to respond gives types below  $\frac{1}{2}$  an incentive to wait: if the other party disclose a collective action that is less favorable, he would then respond by disclosing himself; otherwise, he would

keep quiet. The rate at which types below  $1/2$  disclose at equilibrium is the same as in the original dynamic disclosure game. The reason is that only types below  $1/2$  disclose at a strictly positive time. Therefore such types do not use the opportunity to react, in which case incentives to reveal at  $t$  or at  $t + \varepsilon$  are the same as those in Proposition 7. Establishing uniqueness is technically more involved and details are provided in the supplementary appendix.

Notice that every type below  $\frac{1}{2}$  delays the latest time at which he would disclose, relative to his timing of disclosure in the original dynamic game (where an individual cannot disclose immediately after his rival). Hence, for every compromise rule, the ex-ante expected payoff of an individual is lower in this dynamic game than in the original game discussed above. Again, one sees that more opportunities to speak can in fact be damaging in terms of welfare.

## 6 Beyond the Benchmark

**Knowing No or Multiple Collective Actions** Consider a variant of our model where there is a fixed probability  $p$  that an individual is *not* aware of any collective action, while he is aware of a collective action as before with probability  $1 - p$ . Notice that the first individual's expected net gain of disclosing instead of withholding his type  $x$  (thus conditional on being aware of a collective decision) in the static disclosure game is equal to

$$ENG_1^p(x, \sigma_2) = px + (1 - p)ENG_1(x, \sigma_2),$$

where  $ENG_1$  was defined in equation (2). As explained after Proposition 1, our analysis builds on the fact that  $ENG_1$  is increasing in  $x$  and decreasing in  $\sigma_2$ . Clearly,  $ENG_1^p$  inherits these properties from  $ENG_1$ . Hence, similar arguments imply that there is a unique symmetric BNE, which involves threshold strategies. The equilibrium threshold decreases as  $p$  increases (as it is easy to check that  $ENG_1^p$  is increasing in  $p$ ), but is strictly positive for any  $p < 1$ . Comparisons between compromise rules in terms of the efficiency level they induce are unaffected by the introduction of the probability  $p$ . Similar arguments hold for the dynamic game as well. The expected net gain function

of disclosing now instead of later retains similar properties independently of the value of  $p$ . The threshold above which types reveal right away decreases as  $p$  increases. The rate at which lower types disclose remains the same -  $\frac{(1-2y)f(y)}{\delta y F'(y)}$ . Inefficiency in the form of delay persists for any  $p < 1$ , and comparisons between compromise rules in terms of the efficiency level they induce is unaffected by the introduction of the probability  $p$ .

Consider now a case where each individual knows about the feasibility of  $k$  collective actions whose associated payoffs fall on the line. Formally,  $i$ 's type is a subset  $O_i$  of  $[0, 1]$  (each element of  $O_i$  determines  $i$ 's own payoff, following the same convention as first described in Section 2) that contains  $k$  elements obtained from repeated independent draws that follow the density  $f$ . Suppose that the first individual has disclosed a subset  $X_1$  of  $O_1$ , while the second individual has disclosed a subset  $X_2$  of  $O_2$ . The feasible set in the space of utilities in that case is the smallest triangle that contains all the following vectors:  $(0, 0)$ ,  $(x_1, 1 - x_1)$ , for all  $x_1 \in X_1$ , and  $(1 - x_2, x_2)$ , for all  $x_2 \in X_2$ . A compromise rule is "welfarist" if final utilities depend only on the utility possibility set. Nash's (1950) bargaining model is welfarist, and so are in particular his, Raiffa's, and Kalai and Smorodinsky's solutions. An example of compromise rule that is not welfarist is one that picks the middle option when three have been disclosed. We restrict our discussion to the case of welfarist compromise rules. Hence the compromise depends only on the triangle with extreme points  $(0, 0)$ ,  $(\bar{x}, 1 - \bar{x})$  and  $(1 - \bar{y}, \bar{y})$ , where  $(\bar{x}, 1 - \bar{x})$  is the utility pair that is most advantageous to the first individual among those that are available and  $(1 - \bar{y}, \bar{y})$  is the one that is most advantageous for the second individual.

Consider now the extended static or dynamic disclosure game where a pure strategy for either individual in the static game is any subset of his type, and a pure strategy for either individual in the dynamic game determines which subset of his type to disclose and at which time. It is not difficult to show that individuals' strategies depend only on the collective action that is most advantageous to them given those they are aware of. BNEs of the extended static or dynamic disclosure games are thus isomorphic to the BNEs of the original



static or disclosure games where (single-valued) types are drawn according to the density  $f^k$  (which is the density of the maximum of  $k$  independent draws following  $f$ ). This observation thus allows to extend all our results to cases where parties know multiple collective actions.

**Beyond the Straight Line** Recall that  $X$  represents the set of utility pairs associated with collective decisions that individuals may learn about.<sup>18</sup> Most of our results build on two key features of  $X$ . First it must be symmetric:  $(y, x) \in X$  whenever  $(x, y) \in X$ . Second interests of the two individuals must be strictly opposed: if the first individual prefers  $x$  over  $x'$ , then the second prefers  $x'$  over  $x$ . Any such set  $X$  can be described via a strictly decreasing isomorphism  $g : [0, 1] \rightarrow [0, 1]$  such that  $g(0) = 1$ ,  $g(1) = 0$ , and  $g(1 - x) = x$ , for all  $x \in [0, 1]$ . All our results and associated proofs, except Corollary 1 and Proposition 6, extend to this more general environment simply by replacing  $1 - y$  with  $g^{-1}(y)$  and  $1/2$  by  $u^*$ , where  $u^*$  is the unique fixed point of  $g$  (i.e.  $u^* = g(u^*)$ ). For instance, Proposition 1 holds with the threshold

$$\theta = \sup\{x \in [0, u^*] \mid xF(x) + \int_{y=x}^1 (r_1(x, y) - g^{-1}(y))f(y)dy < 0\}.$$

The unique symmetric BNE in the dynamic disclosure game is given by

$$\tau^*(x) = \begin{cases} 0 & \text{if } x \geq \theta \\ \int_x^\theta \frac{(g^{-1}(y)-y)f(y)}{\delta y F(y)} dy & \text{if } x < \theta, \end{cases}$$

where

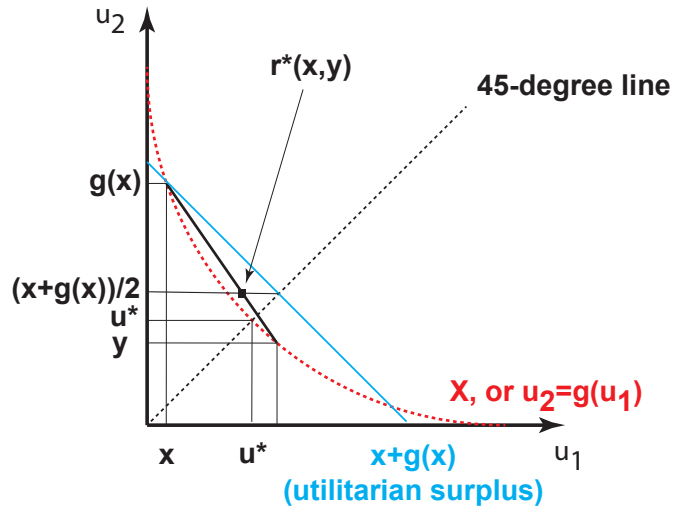
$$\theta = \sup\{x \in [0, u^*] \mid \int_{y=x}^1 (r_1(x, y) - g^{-1}(y))f(y)dy < 0\}.$$

Propositions 4, 5 and 8 on the relative efficiency of different compromise rules hold with the partial ordering from Definition 2 (simply replace  $1/2$  by  $u^*$  in that definition).

On the other hand, considering sets  $X$  different from a straight line has non-trivial implications as far as the optimal solution is concerned (see Corollary 1). First, the Nash solution need not be regular anymore. Second, even if

---

<sup>18</sup>To avoid any possible confusion, we emphasize that  $X$  is not the Pareto frontier of the set of feasible agreements as in a standard bargaining problem. Instead, the set of feasible (efficient) agreements consists of two points randomly drawn from  $X$ .



it is regular, it may be dominated by other regular rules. Characterizing the most efficient regular compromise rule for any symmetric and decreasing  $g$  remains an open question. However, Proposition 5 implies an algorithm that transforms any pair of regular solutions into a regular solution, which is at least as efficient as each of the two original solutions.

More can be said if we further restrict  $g$ . In particular, we are able to characterize the most efficient regular compromise rule if we assume that  $g$  is differentiable and either convex or concave.

We start with the case where  $g$  is differentiable and *convex*. This case is interesting beyond mere generality. It allows to capture situations where an individual would not disclose a collective action which is only slightly worse for him but much better for the other individual (e.g., the first individual may disclose  $(1, 1)$  but not  $(0, 100)$ ), resulting in large efficiency losses ex-ante.

Let  $r^*$  be the compromise rule defined as follows. If two collective actions have been disclosed, it selects the lottery over these two options, which maximizes the expected payoff to the individual in the weaker position, subject to the constraint that the other individual's expected payoff is at least half the utilitarian surplus (see Figure above). This definition retains the idea of favoring the individual in the weaker position within the class of regular compromise rules. If only one collective action has been disclosed,  $r^*$  selects it with certainty. Observe that  $r^*$  coincides with the Nash solution when  $g$  is linear, but not in general.

**Proposition 12** *Suppose that  $g$  is convex. Then  $r_{CF}$ ,  $r_{KS}$ ,  $r_N$  and  $r^*$  are all regular compromise rules. In addition,  $r^*$  is the most efficient regular rule.*

To analyze the case where  $g$  is differentiable and *concave*, we use the following “duality” argument. For any payoff pair  $(u, g(u))$  we define a dual pair  $(v, h(v))$  where  $v \equiv 1 - u$  and  $h(v) \equiv 1 - g(1 - v)$ . It follows that  $h(v)$  is differentiable, decreasing and convex. Let  $r$  be a regular compromise rule defined on the set of disclosable payoffs,  $\{(v, h(v)) : v \in [0, 1]\}$ . Define the “dual solution” to  $r$  as follows: for any pair of disclosed payoff pairs,  $(u, g(u))$  and  $(g^{-1}(u'), u')$ ,

$$d_i(u, u') = 1 - r_i(1 - u, 1 - u')$$

This mapping from the compromise rule  $r$  to its dual rule  $d$  preserves the regularity of the solutions as well as their ranking in terms of efficiency.

**Proposition 13** *(i) If  $r$  is regular, then so is  $d$ , and (ii) for any pair of regular compromise rules,  $(r, r')$  and their dual rules  $(d, d')$ , we have that  $r \succeq r'$  implies  $d \succeq d'$ .*

Let  $d^*$  be the dual of  $r^*$ , the regular compromise rule defined above, which is most efficient when  $g$  is convex. By Proposition 13,  $d^*$  is the most efficient regular compromise rule when  $g$  is concave. Note that the coin-flip rule and both Nash and Kalai-Smorodinsky, defined over a convex  $g$ , admit a dual regular compromise rule when  $g$  is concave. However, apart for the coin-flip rule, their dual does not correspond to the definition of the original rule (e.g., the dual of Nash does not select the payoff pair that maximizes the product of the individuals’ payoffs).

Extending our analysis to cases where the collective action known to one individual may be Pareto inferior to the collective action known to the other individual would require developing entirely new arguments. Our analysis was indeed grounded in the fact that an individual’s expected net gain increases with his type (see discussion following Proposition 1), while this property does not hold, for instance, when types are drawn from the uniform distribution on  $[0, 1]^2$ . Similarly, extending our analysis to non-symmetric distributions of types is left as an open problem for future research.

## References

- Bulow, J. and P. Klemperer. 1999. The Generalized War of Attrition. *American Economic Review* **89**, 175-89.
- Fisher, R., W. Ury, and B. Patton, 1991. Getting to Yes: Negotiating Agreement without Giving In. Houghton Mifflin, New York.
- Frankel, D.M. 1998. Creative Bargaining. *Games and Economic Behavior* **23**, 43-53.
- Gul, F. and R. Lundholm. 1995. Endogenous Timing and the Clustering of Agents' Decisions. *Journal of Political Economy* **103**, 1039-66.
- Hao, L. and W. Suen., 2009. Decision-Making in Committees. *Canadian Journal of Economics* **42**, 359-392.
- Haynes, J. M., 1986. Avoiding Traps Mediators Set for Themselves. *Negotiation Journal* **2**, 187-194.
- Kalai, E., 1977. Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons. *Econometrica* **45**, 1623-1630.
- Kalai, E., and D. Samet, 1985. Monotonic Solutions to General Cooperative Games. *Econometrica* **53**, 307-327.
- Kalai, E., and M. Smorodinsky, 1975. Other Solutions to Nash's Bargaining Problem. *Econometrica* **43**, 513-518.
- Luce, R.D., Raiffa, H., 1957. Games and Decisions. Wiley, New York.
- Milgrom, P. and J. Roberts. 1986. Relying on the Information of Interested Parties. *Rand Journal of Economics* **17**, 18-32.
- Nash J., 1950. The bargaining problem. *Econometrica* **18**, 155-162.
- Ponsati, C. and J. Sákovics. 1995. The War of Attrition with Incomplete Information. *Mathematical Social Sciences* **29**, 239-54.
- Sobel, J., 2010. Giving and Receiving Advice. Mimeo, University of San Diego.

## Appendix

- Lemma 1**
1.  $ENG_i(x, \sigma_{-i})$  is weakly increasing in  $x$ .
  2. If the probability of disclosure is lower than 1 under  $\sigma_{-i}$ , then  $ENG_i(x, \sigma_{-i})$  is strictly increasing in  $x$ .

3. If  $\hat{\sigma}_{-i}(y) \geq \sigma_{-i}(y)$ , for each  $y \in [0, 1]$ , then  $ENG_i(x, \hat{\sigma}_{-i}) \leq ENG_i(x, \sigma_{-i})$ .

Proof: We assume  $i = 1$ . A similar argument applies to the second individual. The fact that it is non-decreasing in  $x$  follows immediately from the monotonicity condition on  $r$ . If  $\{y \in X | \sigma_2(y) < 1\}$  has a strictly positive measure, then  $ENG$  is strictly increasing in  $x$  via the first term in its definition. The third property follows from the fact that  $r_1(x, y) - (1 - y) \leq x$ , for each  $(x, y) \in [0, 1]^2$ , which itself follows from the fact that  $r_1(x, y) \leq \max\{x, 1 - y\}$ , since  $r$  selects a convex combination between  $(x, 1 - x)$  and  $(1 - y, y)$ . ■

**Proof of Proposition 1** For each  $\epsilon \in [0, 1]$ , let  $\sigma_i^\epsilon$  be the “threshold strategy” defined by  $\sigma_i^\epsilon(x) = 0$ , for each  $x < \epsilon$ , and  $\sigma_i^\epsilon(x) = 1$ , for each  $x > \epsilon$ . We prove in the Supplementary Appendix that there exists  $\eta > 0$  such that

$$\forall x \in [0, \eta], \forall \epsilon \in [0, \eta] : ENG_i(x, \sigma_{-i}^\epsilon) < 0. \quad (6)$$

In other words, there exists a positive  $\eta$  small enough such that types below  $\eta$  prefer not to disclose when the other individual follows a threshold strategy whose threshold falls below  $\eta$ .

(EXISTENCE) The first property from Lemma 1 implies that there exists a best response to any strategy, and that any such best response is a threshold strategy: if  $\sigma_i^*$  is a best response against  $\sigma_{-i}$ , then there exists a unique  $\theta_i \in [0, 1]$  such that  $\sigma_i^*(x) = 0$ , for each  $x$  such that  $x < \theta_i$ , and  $\sigma_i^*(x) = 1$ , for each  $x \in [0, 1]$  such that  $x > \theta_i$ . Such a threshold strategy will also be denoted  $\sigma_i^{\theta_i}$ . The existence of a symmetric BNE is thus equivalent to the existence of a fixed point to the correspondence that associates  $i$ 's optimal threshold strategy to each of the opponent's threshold strategies, or  $\theta_i = BR_i(\theta_{-i})$  for short. This will follow from Brouwer's fixed-point theorem after showing that  $BR_i$  is continuous. Let thus  $(\theta(k))_{k \in \mathbb{N}}$  be a sequence of real numbers in  $[0, 1]$  that converges to some  $\theta$ . Suppose on the other hand that  $BR_i(\theta(k))$  converges to some  $\theta' \neq BR_i(\theta)$ . To fix ideas, we'll assume that  $\theta' > BR_i(\theta)$  (a similar reasoning applies if the inequality is reversed). Hence there exists  $K$  such that  $\frac{BR_i(\theta) + \theta'}{2} < BR_i(\theta(k))$ , for all  $k \geq K$ , and

$$ENG_i\left(\frac{BR_i(\theta) + \theta'}{2}, \sigma^{\theta(k)}\right) < 0.$$

Taking the limit on  $k$ , we get

$$ENG_i\left(\frac{BR_i(\theta) + \theta'}{2}, \sigma^\theta\right) \leq 0,$$

by continuity of the integral with respect to its bounds, but which thus leads to a contradiction, since  $\frac{BR_i(\theta) + \theta'}{2} > BR_i(\theta)$ . Hence  $BR_i$  is indeed continuous, and admits a fixed-point.

(UNIQUENESS) Suppose, on the contrary, that one can find two symmetric BNE's. Let  $\theta$  and  $\theta'$  be the two corresponding common thresholds that the two individuals are using. Assume without loss of generality that  $\theta' > \theta$ , and let  $\hat{\theta}$  be a number that falls between  $\theta$  and  $\theta'$ . Lemma 1 and the definition of the thresholds imply  $0 < ENG_1(\hat{\theta}, \sigma_2^\theta) \leq ENG_1(\hat{\theta}, \sigma_2^{\theta'}) < 0$ , which is impossible. This establishes the uniqueness of the symmetric BNE.

(THRESHOLD IN  $]0, 1/2]$ ) We know from equation (6) that the equilibrium threshold must be strictly positive. We now establish that it is no greater than  $1/2$ . First observe that  $\int_{y=1/2}^1 [r_1(\frac{1}{2}, y) - (1 - y)]f(y)dy \geq 0$ , because  $r_1(\frac{1}{2}, y) \geq 1 - y$ , for all  $y \geq 1/2$ . Adding  $1/2F(1/2)$  to this expression leads to a strictly positive number, and hence  $BR_1(1/2) \leq 1/2$ . We also know from equation (6) that  $BR_1(\eta) \geq \eta$ . Hence  $BR_i$  must have its fixed point between  $\eta$  and  $1/2$ .

(CLOSED-FORM FORMULA) We know that the threshold  $\theta$  is given by  $\theta = BR_1(\theta) = \sup\{x \in [0, 1/2] \mid ENG_1(x, \sigma^\theta) < 0\}$ . Lemma 1 implies that  $ENG_1(x, \sigma^x) \leq ENG_1(x, \sigma^\theta)$ , for all  $x < \theta$ , and  $ENG_1(x, \sigma^\theta) \leq ENG_1(x, \sigma^x)$ , for all  $\theta < x$ . Hence  $\theta = \sup\{x \in [0, 1/2] \mid ENG_1(x, \sigma^x) < 0\}$ , which establishes equation (1). ■

**Proof of Proposition 3** Arguments developed in the proof of Proposition 1 imply that any BNE must involve threshold strategies. Hence, a BNE is efficient only if at least one of the two individuals follows a fully revealing strategy. To fix ideas, suppose that we have a BNE in which the second individual systematically reveals the collective action he is aware of. Note that the first individual's expected net gain from disclosing when of type  $\frac{1}{2}$  is

given by

$$\int_{y=0}^1 [r_1(\frac{1}{2}, y) - (1 - y)]f(y)dy.$$

This expression can be decomposed into two components: one where the opponent's type is below  $\frac{1}{2}$ , and another where his type is above  $\frac{1}{2}$ . The second component may be rewritten as follows. First, by anonymity,  $r_1(\frac{1}{2}, y)$  by  $r_2(y, \frac{1}{2})$ . Second, by ex-post efficiency,  $r_2(y, \frac{1}{2}) = 1 - r_1(y, \frac{1}{2})$ . Third, by symmetry of  $r$ ,  $r_1(y, \frac{1}{2}) = r_1(\frac{1}{2}, 1 - y)$ . Making the change of variable  $y' \equiv 1 - y$ , it follows that the net expected gain of type  $\frac{1}{2}$  equals

$$\int_{y=0}^{\frac{1}{2}} [r_1(\frac{1}{2}, y) - (1 - y)]f(y)dy + \int_{y'=0}^{\frac{1}{2}} [1 - r_1(\frac{1}{2}, y') - y']f(1 - y')dy. \quad (7)$$

This expression is strictly positive, since  $f(x) > f(1 - x)$ , for all  $x \geq 1/2$ , and hence the threshold for the first individual's best response strategy is strictly smaller than  $1/2$ . Let's call it  $\theta$ . Then the second individual's expected net gain of revealing when of type  $y = 0$  is equal to

$$\int_{x=\theta}^1 (r_2(x, 0) - (1 - x))f(x)dx.$$

Observe that  $\int_{x=1/2}^1 (r_2(x, 0) - (1 - x))f(x)dx \leq 0$ , since  $r_2(x, 0) \leq (1 - x)$ , for all  $x$ . Also, it must be that  $\int_{x=\theta}^{1/2} (r_2(x, 0) - (1 - x))f(x)dx < 0$  for any regular  $r$ , as  $r_2(x, 0) \leq r_2(0, 0) = 1/2 < 1 - x$ , for all  $0 < x < 1/2$ . Hence systematic revelation cannot be a best response when the other individual uses a threshold strategy that discloses some types below  $1/2$ , and there is no way to achieve efficiency in any BNE of the disclosure game. ■

**Proof of Proposition 4** Recall from Proposition 1 that the unique symmetric BNE of the disclosure game associated with any regular compromise rule involves threshold strategies, whose common threshold falls in the interior of  $[0, \frac{1}{2}]$ . Let  $\theta$  be the threshold associated with  $r$ , and  $\theta'$  be the threshold associated with  $r'$ . Notice that

$$ENG_1^r(\theta', \sigma_2^{\theta'}) \leq 0. \quad (8)$$

Indeed, this inequality actually holds pointwise, since  $r' \succeq r$  and the second individual withholds his information when  $y < \theta'$ , and is thus preserved through summation.

We now conclude the proof by showing that  $\theta \geq \theta'$ . Suppose, on the contrary, that  $\theta < \theta'$ , and let  $\hat{\theta}$  be a number that falls between  $\theta$  and  $\theta'$ . Remember from Lemma 1 that an individual's expected net gain is increasing in his own type. Inequality (8) thus implies that  $ENG_1^r(\hat{\theta}, \sigma_2^{\theta'}) < 0$ . Remember also from Lemma 1 that an individual's expected net gain does not increase when the opponent reveals more, and hence  $ENG_1^r(\hat{\theta}, \sigma_2^{\theta}) < 0$ . This contradicts the fact that the threshold strategies associated with  $\theta$  form a BNE of the disclosure game associated with  $r$  (as it should be optimal for the first individual to reveal his type  $\hat{\theta}$  since it is larger than  $\theta$ ). ■

**Proof of Corollary 1** : Let  $x$  be a number smaller or equal to  $1/2$ . We must prove that  $r_N(x, y)$  is more advantageous to the first individual than  $r(x, y)$ , for all  $y \geq x$ . This is obvious when  $y \geq 1/2$  since the Nash solution picks the right-most option in that region. Since  $r$  is anonymous, it must be that  $r(x, x) = (1/2, 1/2)$ . Monotonicity implies that  $r_1(x, y) \geq 1/2$  for each  $y \in [x, 1/2]$ , hence, the desired inequality when compared to the Nash solution which always picks  $1/2$  in that region. ■

**Proof of Proposition 5** To establish (i), first notice that  $r \vee r'$  and  $r \wedge r'$  are well-defined, as  $\min\{x, 1 - y\} = \min\{1 - x, y\}$  if and only if  $x = y$ , in which case  $r_1(x, y) = r'_1(x, y) = \frac{1}{2}$ . Next, it is easy to check that  $r \vee r'$  and  $r \wedge r'$  satisfy efficiency and symmetry. Next, consider a pair of real-valued functions,  $\phi(\cdot)$  and  $\varphi(\cdot)$ , defined over some subset of  $\mathbb{R}$ . If both  $\phi(\cdot)$  and  $\varphi(\cdot)$  are non-decreasing then so are  $\max\{\phi(\cdot), \varphi(\cdot)\}$  and  $\min\{\phi(\cdot), \varphi(\cdot)\}$ . Also, if  $\alpha$  is a real number such that  $\phi(\alpha) = \varphi(\alpha)$ , then  $h(\cdot)$ , where  $h(z) = \phi(z)$  if  $z \leq \alpha$  and  $h(z) = \varphi(z)$  if  $z \geq \alpha$ , is also non-decreasing. Hence  $r \vee r'$  and  $r \wedge r'$  is monotone (apply these simple facts to  $x$  and  $y$  in turn). We conclude by establishing (ii) and (iii). For any  $x \leq 1/2$  and  $y \geq x$ , we have that  $\min\{x, 1 - y\} \leq \min\{1 - x, y\}$ , in which case  $(r \vee r')_1(x, y)$  is equal to  $\max\{r_1(x, y), r'_1(x, y)\}$ . By the definition of the partial order  $\succeq$ , it follows that



$r \vee r' \succeq r$  and  $r \vee r' \succeq r'$ . A similar argument implies that  $r \succeq r \wedge r'$  and  $r' \succeq r \wedge r'$ . ■

**Proof of Proposition 6** If  $f$  is symmetric, then expression (7) is equal to zero, and hence the first individual's best response to the fully revealing strategy is to reveal if and only if his type is larger or equal to  $1/2$ . Let's check now that systematic revelation is a best response for the second individual. Notice that

$$\int_{x=1/2}^1 (r_2(x, 0) - (1 - x))f(x)dx$$

cannot be strictly negative, as the second individual's expected net gain of disclosing would then be negative when his type is very small. Given that  $r_2(x, 0) \leq 1 - x$ , for all  $x \in [1/2, 1]$  and that  $r_2$  is continuous, it must thus be that  $r_2(x, 0) = 1 - x$ , for all  $x \in [1/2, 1]$ . The third regularity condition implies that  $r_2(x, y) = 1 - x$ , for all  $y \leq 1/2$  and all  $x \in [1/2, 1 - y]$ . The second regularity condition then implies that  $r_2(x, y) = \max\{1 - x, y\}$  if  $\max\{1 - x, y\} \leq 1/2$  and  $= \min\{1 - x, y\}$  if  $\min\{1 - x, y\} \geq 1/2$ . Similar conditions apply for the first individual's payoffs, by anonymity. Consider now a case where both  $x$  and  $y$  are no larger than  $1/2$ . The third regularity condition implies that the first individual's payoff is no larger than  $r_1(1/2, y) = 1/2$  and the second individual's payoff is no larger than  $r_2(x, 1/2) = 1/2$ . Hence  $r(x, y) = (1/2, 1/2)$ . Anonymity implies that a similar argument applies when both  $x$  and  $y$  are no smaller than  $1/2$ . Hence  $r$  must coincide with the Nash solution. When  $r$  is the Nash solution, the second individual is indifferent between revealing or not when of type  $y = 0$  given that the opponent systematically reveals. Hence full disclosure is a best response and we have identified an efficient BNE. ■

**Proof of Proposition 7 (Existence)** We prove that the strategy  $\tau^*$  is indeed part of a symmetric BNE. We start by showing that reporting at  $\tau^*(x)$  is optimal, for any  $x \in [0, \theta[$ . Consider first the possibility of revealing at positive times. The function  $\tau^*$  being invertible on  $[0, \theta[$ , we can identify any positive time with the type speaking at that time. The expected utility from

revealing at  $\tau^*(z)$  when of type  $x$  is equal to

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

for each  $z \in [0, \theta]$ . This expression is differentiable, and the derivatives is equal to

$$xf(z)e^{-\delta\tau^*(z)} - \delta x(\tau^*)'(z)F(z)e^{-\delta\tau^*(z)} - (1-z)f(z)e^{-\delta\tau^*(z)},$$

or

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

after rearranging the terms and using the definition of  $\tau^*$  to compute  $(\tau^*)'$ . 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  $\tau^*(x)$ , when of type  $x$ . Deviating to report at zero is not profitable either, as the expected payoff in that case is  $xF(\theta) + \int_{y=\theta}^1 r_1(x, y)f(y)dy$ , which is equal to  $U(\theta|x) + \int_{y=\theta}^1 (r_1(x, y) - (1-y))f(y)dy$ . For any  $\epsilon > 0$  small enough, using the third regularity condition, this last expression is lower or equal to

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

The second term is negative, for all  $\epsilon > 0$ , by definition of  $\theta$ . Hence, taking the limit when  $\epsilon$  decreases to zero, we get that the expected utility of reporting at zero is no greater than  $U(\theta|x)$ , which in turn, by our previous reasoning, is smaller than the expected utility of reporting at  $\tau(x)$ . This establishes the optimality of  $\tau^*$ , for any type strictly in between 0 and  $\theta$ .

Consider now a type  $x \in ]\theta, 1]$ . The expected utility of revealing at a time  $t$  is equal to  $U(z|x)$ , where  $z$  is the unique real number in  $[0, \theta[$  such that  $\tau^*(z) = t$ . Our earlier reasoning regarding  $U$ 's derivative implies that this expected utility is strictly lower than  $U(\theta|x)$  (since  $z < \theta \leq x$ ), which is equal to  $xF(\theta) + \int_{y=\theta}^1 (1-y)f(y)dy$ . Notice that  $\int_{y=\theta}^{\theta+\epsilon} (r_1(x, y) - (1-y))f(y)dy$  converges to zero as  $\epsilon$  decreases to zero. Hence  $U(z|x) < U(\theta|x) + \int_{y=\theta}^{\theta+\epsilon} (r_1(x, y) - (1-y))f(y)dy$ ,

for any  $\epsilon > 0$  that is small enough. This in turn implies that

$$U(z|x) < U(\theta|x) + \int_{y=\theta}^{\theta+\epsilon} (r_1(x, y) - (1-y))f(y)dy + \int_{y=\theta+\epsilon}^1 (r_1(\theta+\epsilon, y) - (1-y))f(y)dy,$$

by definition of  $\theta$ . Applying now the third regularity condition ( $\epsilon$  is small enough so that  $x > \theta + \epsilon$ ), we conclude that

$$U(z|x) < U(\theta|x) + \int_{y=\theta}^{\theta+\epsilon} (r_1(x, y) - (1-y))f(y)dy + \int_{y=\theta+\epsilon}^1 (r_1(x, y) - (1-y))f(y)dy.$$

The right-hand side is the expected utility for type  $x$  of revealing at zero, and we have thus proved the optimality of  $\tau^*$  for any type no smaller than  $\theta$ .

Finally, if  $x = \theta$ , then a similar reasoning as in the last paragraph implies that revealing at a positive time leads to a payoff that is no larger than  $U(\theta|\theta) = \theta F(\theta) + \int_{y=\theta}^1 (1-y)f(y)dy$ . Revealing at zero leads to the expected payoff  $\theta F(\theta) + \int_{y=\theta}^1 r_1(x, y)f(y)dy$ . The condition  $\int_{y=\theta}^1 (r_1(\theta, y) - (1-y))f(y)dy \geq 0$  thus guarantees that revealing at zero is optimal for type  $\theta$ , as desired. ■

**Proof of Proposition 7 (Uniqueness)** Let  $r$  be a regular compromise rule that satisfies condition (3), and let  $\tau$  be a strategy that is part of a symmetric BNE in the original dynamic game. We have to show that  $\tau = \tau^*$ . We proceed in various steps.

**Step 1** *The probability of having a type disclose at any given strictly positive time or at infinity (i.e. not disclose at all) is zero. Formally,  $\int_{y \in \tau^{-1}(\infty)} f(y)dy = 0$  and  $\int_{y \in \tau^{-1}(t)} f(y)dy = 0$ , for all  $t > 0$ .*

Proof: See Supplementary Appendix. □

**Step 2**  *$\tau$  is strictly decreasing with respect to time: if  $x' > x$  and  $\tau(x) > 0$ , then  $\tau(x') < \tau(x)$ ; if  $x' > x$  and  $\tau(x) = 0$ , then  $\tau(x') = 0$ .*

Proof: Let  $x, x' \in [0, 1]$  such that  $x' > x$  and  $\int_{y \in [0, 1] \text{ s.t. } \tau(y) > \tau(x)} f(y)dy > 0$ .

By definition of BNE,  $ENG_1(\tau(x'), \tau(x), x') \geq 0$ . If  $\tau(x') > \tau(x)$ , then<sup>19</sup>  $ENG_1(\tau(x'), \tau(x), x) > 0$ , which contradicts the optimality of reporting at  $\tau(x)$  when of type  $x$ . Hence, it must be that  $\tau(x') \leq \tau(x)$ .

<sup>19</sup>The second term in the definition of the expected net gain, as stated before the statement of this proposition, is zero, by Step 1.

Suppose now that  $\tau(x) > 0$ . We know from the previous paragraph that  $\tau(x'') \leq \tau(x)$ , for all  $x'' \in ]x, x'[$ . Step 1 implies that there exists  $x'' \in ]x, x'[$  such that  $\tau(x'') < \tau(x)$ . The reasoning from the previous paragraph implies that  $\tau(x') \leq \tau(x'')$ , and hence  $\tau(x') < \tau(x)$ .

We have thus established the two desired properties, but under the assumption that  $\int_{y \in [0,1] \text{ s.t. } \tau(y) > \tau(x)} f(y)dy > 0$ . We now show that this inequality must in fact hold for any  $x > 0$ . Suppose first that  $x$  is such that  $\tau(x) = 0$ . If the inequality does not hold, then it means that the opponent will reveal his type with probability 1 at time 0. Then it is easy to check that  $\int_{y \in [0,1]} r_1(x, y) f(y)dy < \int_{y \in [0,1]} (1 - y) f(y)dy$ , for any  $x \in [0, 1]$  that is small enough. A reasoning similar to the one developed in the second paragraph of the proof of Step 1 (see Supplementary Appendix) would imply a contradiction, namely that a slight delay is a profitable deviation for any such  $x$ . Consider now an  $x$  such that  $\tau(x) > 0$ , let  $t^* = \inf_{y \in [0,x]} \tau(y)$ , and let  $(x_k)_{k \in \mathbb{N}}$  be a sequence in  $[0, x]$  such that  $(\tau(x_k))_{k \in \mathbb{N}}$  decreases towards  $t^*$  as  $k$  tends to infinity. Since  $\tau$  is measurable, we have:

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

Notice that the right-most expression must be strictly positive. We just proved this if  $t^* = 0$ , while, if  $t^* > 0$ , then the opponent does not speak before  $t^*$  if his type is no greater than  $x$ , and the probability of him speaking at  $t^*$  is zero, by Step 1. Hence there exists  $K \in \mathbb{N}$  such that  $\int_{y \in [0,1] \text{ s.t. } \tau(y) > \tau(x_k)} f(y)dy > 0$ , for all  $k \geq K$ . The result from the previous paragraph implies that  $\tau(x) \leq \tau(x_k)$ , for all such  $k$ 's, and hence  $\tau(x) = t^*$ , and  $\int_{y \in [0,1] \text{ s.t. } \tau(y) > \tau(x)} f(y)dy > 0$ , as desired.  $\square$

**Step 3** *Let  $\alpha = \inf\{x \in [0, 1] | \tau(x) = 0\}$ . Then  $\tau$  is continuous on  $]0, \alpha[$ , and  $\lim_{x \rightarrow \alpha^-} \tau(x) = 0$ .*

Proof: Let  $x \in ]0, \alpha[$ , and let  $(x_k)_{k \in \mathbb{N}}$  be a sequence in  $[0, x]$  that converges to  $x$ . Step 2 implies that  $\tau(x_k) \geq \tau(x)$ , for all  $k \in \mathbb{N}$ . Suppose, to the contrary of what we want to prove, that there exists  $\eta > 0$  and  $K \in \mathbb{N}$  such that  $\tau(x_k) > \tau(x) + \eta$ , for all  $k \geq K$ . This implies that no type reveals after  $\tau(x)$

and before  $\tau(x) + \eta$ . Indeed, suppose on the contrary that there exists  $y$  such that  $\tau(y) \in ]\tau(x), \tau(x) + \eta[$ . Step 2 implies that  $y$  is strictly smaller than  $x$ , and hence there exists  $k \geq K$  such that  $y < x_k < x$ . Step 2 implies that  $\tau(x_k) < \tau(y) < \tau(x) + \eta$ , which contradicts the definition of  $K$ . Consider now a type  $y$  for which  $\tau(y)$  is very close to the  $\inf\{\tau(z) | \tau(z) \geq \tau(x) + \eta\}$  (i.e.  $y$  is smaller than  $x$ , but very close to it). Then revealing a bit earlier, let's say at  $\tau(x) + \frac{\eta}{2}$  instead of  $\tau(y)$ , is a profitable deviation since the loss, coming from the opponent's types between  $y$  and  $x$ , can be made as small as needed, while the gain is larger than the gain from getting  $y$  earlier by at least  $\eta/2$  units of time for all the opponent's type who reveal after  $\tau(y)$  ( $y$  is strictly positive if close enough to  $x$ , and so there is a positive probability that the opponent reveals after  $\tau(y)$ ). This contradicts the optimality of revealing  $y$  at  $\tau(y)$ , and hence we have established the left-continuity on  $]0, \alpha[$ , and that  $\lim_{x \rightarrow \alpha^-} \tau(x) = 0$ . A similar reasoning applies to show the right-continuity on  $]0, \alpha[$ .  $\square$

**Step 4**  $\tau(x) = 0$  if and only if  $x \in [\theta, 1]$ , where

$$\theta = \sup\{x \in [0, 1/2] | \int_{y=x}^1 (r_1(x, y) - (1 - y))f(y)dy < 0\}.$$

Proof: Observe first that the function  $g : [0, 1/2] \rightarrow \mathbb{R}$  that associates  $\int_x^1 (r_1(x, y) - (1 - y))f(y)dy$ , to any  $x \in [0, 1/2]$ , is strictly increasing. Suppose that  $x' > x$ . We have:

$$\begin{aligned} g(x') &= \int_{y=x'}^1 (r_1(x', y) - (1 - y))f(y)dy \geq \int_{y=x'}^1 (r_1(x, y) - (1 - y))f(y)dy \\ &> \int_{y=x}^1 (r_1(x, y) - (1 - y))f(y)dy = g(x). \end{aligned}$$

The weak inequality follows from the third regularity condition, while the strict inequality follows from the fact that  $r_1(x, y) - (1 - y) < 0$ , for each  $y \in ]x, x'[$ , as  $1 - y > 1/2$  and  $r_1(x, y) \leq 1/2$  (as a consequence of the second and third regularity conditions), for all such  $y$ 's. Notice also that  $g(0) < 0$ . Indeed,  $r_1(0, y) \leq 1 - y$ , for all  $y \in [1/2, 1]$ , by the first regularity condition, and  $r_1(0, y) \leq 1/2 < 1 - y$ , for all  $y \in [0, 1/2[$ , by the second and third

regularity conditions. Notice finally that  $g(1/2) \geq 0$ , as  $r_1(1/2, y) \geq 1 - y$ , for each  $y \in [1/2, 1]$ , by the first regularity condition. Hence  $\theta$  is well-defined,  $g(x) < 0$ , for each  $x \in [0, 1/2]$  such that  $x < \theta$ , and  $g(x) > 0$ , for each  $x \in [0, 1/2]$  such that  $x > \theta$ .

We now prove that  $\tau(x) > 0$ , for each  $x < \theta$ . Otherwise, there exists  $x < \theta$  such that  $\tau(x) = 0$ . Then  $g(x) < 0$ , and hence

$$\int_{y=\alpha}^1 r_1(x, y) < \int_{y=\alpha}^1 (1 - y)f(y)dy,$$

where  $\alpha = \inf\{y \in [0, 1] | \tau(y) = 0\}$ , because  $r_1(x, y) \leq 1 - y$ , for each  $y \in [\alpha, x]$ , by the first regularity condition. A reasoning similar to the one we did in the second paragraph proof of Step 1 (see Supplementary Appendix) implies that an individual of type  $x$  can improve his payoff by reporting at some small positive time rather than at zero, thereby contradicting the optimality of  $\tau$ . Hence  $\tau(x) > 0$ , for each  $x < \theta$ , as desired.

We now prove that  $\tau(x) = 0$ , for each  $x > \theta$ . First notice that  $\tau(x) = 0$ , for each  $x > 1/2$ . Suppose, on the contrary, that  $\tau(x) > 0$ , for some  $x > 1/2$ . The expected net gain of reporting at 0 instead is strictly positive, as  $r_1(x, y) - (1 - y) \geq 0$ , for all the opponent's types  $y$  that report at 0, and  $x > 1 - y$ , for all the opponent's types  $y > x$  that report at a positive time lower than  $\tau(x)$ . So  $\tau(x) = 0$ , for each  $x > 1/2$ , and we have proved the statement for  $\theta = 1/2$ . Suppose now that  $\theta < 1/2$ . As before, let  $\alpha = \inf\{y \in [0, 1] | \tau(y) = 0\}$ . We know that  $\alpha \leq 1/2$ . Suppose, to the contrary of what we want to prove, that  $\alpha > \theta$ . Let then  $x$  be smaller than  $\alpha$ , but very close to it. Hence  $\tau(x) > 0$ . The expected net gain of revealing at zero instead is equal to:

$$\int_{y=\alpha}^1 (r_1(x, y) - (1 - y))f(y)dy + \int_{y=x}^{\alpha} (x - e^{-\delta\tau(y)}(1 - y))f(y)dy + x(1 - e^{-\delta\tau(x)}) \int_{y=0}^x f(y)dy,$$

which is greater or equal to

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

which is equal to

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

Notice that the first term is  $g(x)$ , which is strictly positive if  $x > \theta$ , and increasing with  $x$ . The second term, on the other hand, can be made as small as desired, by choosing  $x$  large enough, so as to be as closed as needed to  $\alpha$ . Hence the expected net gain for such a type to reveal at zero is strictly positive, which contradicts the optimality of  $\tau$ . This concludes the proof that  $\tau(x) = 0$ , for each  $x > \theta$ .

Finally, we prove that  $\tau(\theta) = 0$ . We have proved that  $\theta = \alpha$ . If  $\tau(\theta) > 0$ , then  $\tau(x) \geq \tau(\theta)$ , for all  $x < \alpha$ , by Step 2, and  $\lim_{x \rightarrow \alpha^-} \tau(x) > 0$ , which would contradict Step 3. Hence  $\tau(\theta) = 0$ , and we are done proving Step 4.  $\square$

**Step 5**  $\tau$  is differentiable on  $]0, \theta[$ , and  $\tau'(x) = \frac{(1-2x)f(x)}{\delta x F(x)}$ , for all  $x \in ]0, \theta[$ .

Proof: Let  $x \in ]0, \theta[$ . The expected net gain of revealing at  $\tau(x + \epsilon)$  instead of  $\tau(x)$  is equal to:

$$\int_{y=x}^{x+\epsilon} (xe^{-\delta\tau(x+\epsilon)} - (1-y)e^{-\delta\tau(y)})f(y)dy + x(e^{-\delta\tau(x+\epsilon)} - e^{-\delta\tau(x)}) \int_{y=0}^x f(y)dy,$$

which is also equal to

$$- \int_{y=x}^{x+\epsilon} (1-y)e^{-\delta\tau(y)}f(y)dy + x(e^{-\delta\tau(x+\epsilon)}F(x+\epsilon) - e^{-\delta\tau(x)}F(x)).$$

In order for  $\tau$  to be optimal, it must be that this expression is non-positive. Dividing by  $\epsilon$ , and taking the limit when  $\epsilon$  decreases to 0, we get:

$$-e^{-\delta\tau(x)}(1-2x)f(x) - x\delta \lim_{\epsilon \rightarrow 0_+} \left[ \frac{\tau(x+\epsilon) - \tau(x)}{\epsilon} \right] e^{-\delta\tau(x)}F(x) \leq 0.$$

A similar reasoning applied to the case that type  $x + \epsilon$  is not better off by reporting at  $\tau(x)$  gives

$$e^{-\delta\tau(x)}(1-2x)f(x) + x\delta \lim_{\epsilon \rightarrow 0_+} \left[ \frac{\tau(x+\epsilon) - \tau(x)}{\epsilon} \right] e^{-\delta\tau(x)}F(x) \leq 0.$$

Combining the two previous inequalities, we conclude that

$$\lim_{\epsilon \rightarrow 0_+} \left[ \frac{\tau(x + \epsilon) - \tau(x)}{\epsilon} \right] = -\frac{(1 - 2x)f(x)}{\delta x F(x)}.$$

A similar reasoning with  $\epsilon < 0$  implies that

$$\lim_{\epsilon \rightarrow 0_-} \left[ \frac{\tau(x + \epsilon) - \tau(x)}{\epsilon} \right] = -\frac{(1 - 2x)f(x)}{\delta x F(x)},$$

which concludes the proof of this step.  $\square$

**Step 6**  $\tau = \tau^*$ .

Proof: Step 4 establishes that  $\tau = \tau^*$  on  $[\theta, 1]$ . Step 5 implies that  $\tau = C + \tau^*$  on  $[0, \theta[$ , for some real number  $C$ . The fact that  $\lim_{x \rightarrow \theta_-} \theta(x) = 0$ , implies that  $C = 0$ , and establishes that  $\tau = \tau^*$  on  $[0, 1]$ .  $\blacksquare$

**Proof of Proposition 8** Given the characterization of the symmetric BNE in Proposition 7, we see that proving  $\tau'(x) \leq \tau(x)$ , for each  $x \in [0, 1]$ , is equivalent to proving  $\theta' \leq \theta$ , where  $\theta$  and  $\theta'$  are the thresholds defined in (4) for  $r$  and  $r'$  respectively. Suppose, to the contrary of what we want to prove, that  $\theta' > \theta$ . Then for any  $\epsilon > 0$  small enough so that  $\theta' - \epsilon > \theta$ , we have  $\int_{y=\theta'-\epsilon}^1 (r_1(\theta' - \epsilon, y) - (1 - y))f(y)dy \geq 0$ , by definition of  $\theta$ . Since  $r' \geq r$ , we must also have  $\int_{y=\theta'-\epsilon}^1 (r'_1(\theta' - \epsilon, y) - (1 - y))f(y)dy \geq 0$ , but this contradicts the definition of  $\theta'$ . Hence  $\theta' \leq \theta$ , as desired.  $\blacksquare$

**Proof of Proposition 9** Assume  $\theta_S > \theta_D$  and let  $\hat{\theta}$  be a type between  $\theta_D$  and  $\theta_S$ . Consider the static disclosure game first. Assume individual  $j$  uses the symmetric equilibrium strategy associated with the threshold  $\theta_S$ . Then for type  $x$  of individual  $i$ , the expected net gain from disclosing, given by  $x F(\theta_S) + \int_{\theta_S}^1 [r_i(x, y) - (1 - y)]f(y)dy$  is positive for all  $x > \theta_S$  and negative for all  $x < \theta_S$ . In particular, it is negative for  $x = \hat{\theta} < \theta_S$ . Since  $\hat{\theta} F(\theta_S)$  is strictly positive, it follows that  $\int_{\theta_S}^1 [r_i(\hat{\theta}, y) - (1 - y)]f(y)dy < 0$ . Because  $\theta_S \leq \frac{1}{2}$ , we have that  $1 - y > \hat{\theta}$  for all  $\hat{\theta} \leq y \leq \theta_S$ . Hence,  $r_i(\hat{\theta}, y) \leq (1 - y)$  for all  $\hat{\theta} \leq y \leq \theta_S$ . Therefore,  $\int_{\hat{\theta}}^1 [r_i(\hat{\theta}, y) - (1 - y)]f(y)dy < 0$ . This contradicts the definition of  $\theta_D < \hat{\theta}$  in Proposition 7.  $\blacksquare$



**Proof of Proposition 10** Observe that the ex-ante expected sum of payoffs is equal to

$$1 - \theta_S^2 \tag{9}$$

since the sum of the individuals' payoffs equals 1 when at least one of them discloses his collective action, and 0 otherwise. Since both individuals are ex-ante symmetric, the ex-ante expected payoff of each is equal to  $(1 - \theta_S^2)/2$ .

A similar reasoning implies that the sum of individuals' ex-ante expected payoffs in the symmetric BNE of the dynamic game is equal to

$$1 - \theta_D^2 + \int_{x=0}^{\theta_D} \int_{y=0}^{\theta_D} e^{-\delta\tau(\max\{x,y\})} dx dy \tag{10}$$

where

$$\tau(x) = \int_x^{\theta_D} \frac{1 - 2y}{\delta y^2} dy = -\frac{1}{\delta\theta_D} + \frac{1}{\delta x} - \frac{2}{\delta} \ln \theta_D + \frac{2}{\delta} \ln x \tag{11}$$

for  $x \leq \theta_D$  and uniform  $f$ . We show in the Supplementary Appendix that (10) is then equal to

$$1 - \theta_D^2 - 2e^{1/\theta_D} \cdot \theta_D^2 \cdot E_i\left(-\frac{1}{\theta_D}\right). \tag{12}$$

Again, symmetry implies that the each individual's ex-ante expected payoff is 1/2 of this expression. Note it does not depend on the discount factor.

Substituting into (9) and (12) the equilibrium thresholds computed in the first Section of the Supplementary Appendix yields the ex-ante expected payoffs for the coin-flip rule: 0.472 in the dynamic game and 0.474 in the static game; for the Kalai-Smorodinsky solution: 0.473 in the dynamic game and 0.476 in the static game; for the Nash solution: 0.481 in the dynamic game and 0.483 in the static game. We see that an individual's ex-ante expected payoff is higher than his expected payoff in the dynamic game for  $r_{CF}$ ,  $r_{KS}$  and  $r_N$ . ■

**Proof of Proposition 11** The proof methodology is comparable to that of Proposition 7, and is thus relegated to the Supplementary Appendix. ■

**Proof of Proposition 12** The proof that  $r_{CF}$ ,  $r_{KS}$ ,  $r_N$  and  $r^*$  are regular can be found in the Supplementary Appendix.

We now focus on the proof that  $r^*$  is optimal. As pointed out at the beginning of Section 6.2, Proposition 4 is one of the results that carry over to any

symmetric  $X$  with no Pareto comparisons. Hence Proposition 12 will follow after showing that  $r^* \succeq r$ , for all regular compromise rule  $r$ , which amounts to show  $r_1^*(x, y) \geq r_1(x, y)$ , for all  $x \leq u^*$  and all  $y \geq x$ , where  $u^*$  is the real number such that  $u^* = g(u^*)$ . We may also assume without loss of generality that  $y \leq g(x)$ , as otherwise our argument applies by renaming  $(x, g(x))$   $(g^{-1}(y), y)$ , and vice-versa. We will be done after showing that monotonicity on  $r$  implies that  $r_2(x, y)$  is no smaller than half the utilitarian surplus (since  $r_1^*(x, y)$  is the first individual's largest feasible payoff under that constraint). The utilitarian surplus is achieved at  $(x, g(x))$ , since  $g$  is convex. Changing  $(g^{-1}(y), y)$  into  $(g^{-1}(x), x)$  does not increase the second individual's payoff (since  $x \leq y$ ), while the second individual's payoff become equal to half the utilitarian surplus of the original problem (the new problem being solved by symmetry). ■

**Proof of Proposition 13** (i) By construction, the dual solution  $d$  is anonymous and ex-post efficient. To establish monotonicity, suppose we move from the payoff pair  $(u, g(u))$  and  $(g^{-1}(u'), u')$  to  $(u^*, g(u^*))$  and  $(g^{-1}(u'), u')$ . If  $u^* > u$ , then  $1 - u^* < 1 - u$ . Then  $r_i(1 - u^*, 1 - u') \leq r_i(1 - u, 1 - u')$ , because  $r$  is monotone. Hence,  $d_i(u^*, u') \geq d_i(u, u')$ . Essentially the same argument applies if we were to change  $(g^{-1}(u'), u')$  holding fixed  $(u, g(u))$ .

(ii) Define  $\phi$  as the value in  $[0, 1]$  that satisfies  $\phi = g(\phi)$ . We have to show that  $d_1'(u, u') \geq d_1(u, u')$ , for all  $u \leq \min\{\phi, u'\}$ . By definition of  $d$ , this is equivalent to showing that  $r_1'(v, v') \leq r_1(v, v')$ , where  $v := 1 - u$  and  $v' := 1 - u'$ . Since  $r(v, v')$  and  $r'(v, v')$  belong to the same segment with negative slope, this is equivalent to  $r_2'(v, v') \geq r_2(v, v')$ , which follows from the fact that  $r' \succeq r$ . ■