Rules With Discretion and Local Information

Renee Bowen, David Kreps, and Andy Skrzypacz

December 2012

Abstract: To ensure that individual actors take certain actions, community enforcement may be required. This can present a rules-versus-discretion dilemma: It can become impossible to employ discretion based on information that is not widely held, because the wider community is unable to tell whether the information was used correctly. Instead, actions may need to conform to simple and widely verifiable rules. We study when discretion in the form of exceptions to the simple rule can be implemented, if the information is shared by the action taker and a second party, who is able to verify for the larger group that an exception is warranted. In particular, we compare protocols where the second party excuses the action taker from taking the action ex ante with protocols where the second party instead forgives a rule-breaking actor ex post, finding that the latter is, in general, useful in a wider variety of circumstances.

1. Introduction

Economists have long advanced the notion that, in some contexts, decision makers should forbear from using their best judgement and instead follow some rule. The context of central banking is perhaps the first in which this idea was explored, by Simons (1936) and subsequently (for instance) by Modigliani (1977) and Lucas (1980). Other prominent contexts include financial accounting (Bratton [2003]; Schipper [2003]; Barth [2008]), the practice of medicine (Kessler [2011]), and jurisprudence (Freed [1992]; Becker [1997]).

The argument for the use of discretion typically comes down to (superior) information held by the decision maker; the counter-arguments in favor of the application of a rule vary, but among them (and relevant to this paper) are the need for ex post verifiability by a wider group of individuals that the decision taken was “appropriate.” In particular, where one depends on community enforcement of decisions made by individuals, it may be necessary that individuals in the community can perform this ex post verification.

While much of the literature has stressed a dichotomous choice between rules and discretion, it will come as no surprise to an observer of the real world that the matter is more nuanced: Decision makers are given limited amounts of discretion in some cases; in other cases rules are established, but exceptions are permitted and/or violations to the rule are forgiven.

For instance, Bowen (2011) examines the decision-making and enforcement procedures of the General Agreement on Tariffs and Trade (GATT). (For related work, see also Maggi [1999].) Under GATT, general rules are applied on a global basis to the trading relations between pairs of countries that, broadly speaking,

---

1 Comments by referees and editors, and seminar participants at UC Santa Barbara, Stanford Graduate School of Business, the NBER Organization Economics Meeting, and the NBER Macroeconomics Within and Across Borders Meetings are gratefully acknowledged, as is the financial support of the Stanford Graduate School of Business.
say that departures from free trade should be punished (with a retaliatory tariff). However, as Bowen
observes, numerous departures from free trade go unpunished. That is, country A may exercise some
discretion in applying a retaliatory tariff to country B when country A sympathizes with the reason country
B departed from free trade. Such an equilibrium may not be possible with only bilateral relationships,
because it is very costly to withhold tariff retaliation. But the threat of the GATT agreement falling apart
(multilaterally) provides enough incentive to do this.

Or consider the case of the Toyota system of subcontracting, in which Toyota forms long-term
“strategic partnerships” with firms that supply Toyota with major sub-assemblies (see Milgrom and
Roberts [1992]; Kreps [2004, Chapter 24]). Toyota, acting as a hierarchical superior, more or less
dictates terms (prices to be paid, quantities to be ordered, designs to be followed) to its strategic partners,
who moreover are expected to (and do) make sunk-cost investments in the relationship. Having made
these sunk-cost investments, the suppliers are protected from hold-up exploitation by Toyota largely
through the threat of the collective action of the all of Toyota’s suppliers, to be triggered if any one of
them is unfairly treated by Toyota. Faced with the collective power of its suppliers, Toyota has a powerful
incentive to maintain its reputation for fair-dealing with individual suppliers. Toyota follows general
rules, so that suppliers can verify that Toyota has dealt fairly in individual cases. But exceptions to these
rules are sometimes made, and Toyota expects the affected supplier to reassure its peers that the exception
was legitimate; Toyota even provides the forum—various industry groups of its suppliers—at which such
reassurances can be offered.

Or think of the case of the faculty of a school or department. Rules are adopted for how individual
faculty members behave, what are their responsibilities, and so forth. But deans and department chairs,
acting on behalf of the school or department, will sometimes allow exceptions to or forgive transgressions
against these rules in specific cases, expecting that members of the department/school will trust that the
exceptions/forgiveness are allowed for appropriate reasons.

We contend that in all these and in similar cases, a key is that the “private information” on which basis
the exception is made is not private in the strict sense that only one party (the party who does not follow
the rule) has access to the information, but instead is local, by which we mean: held by both the active
party who does not follow the rule and by the second party whose (lack of) action reassures members of
the larger community that a legitimate exception was made. So, for instance, we expect that exceptions
to rules are likely to be more prevalent in circumstances where the second party is more likely to be
able to judge if an exception was appropriate; i.e., where information is local rather than private. Casual
empiricism suggests that this prediction is true; certainly it seems the case that, in the context of GATT
enforcement, departures from free trade are more likely to go unpunished the closer the two countries are
diplomatically, which presumably means the more likely it is that the offended party will understand why
the offending party did what it did.

2 Our use of the adjective “local” echoes the way the term is used in Wolitzky (2012). But Wolitzky is concerned with local
information about actions the players take, while our concern is with local information about payoff-relevant but exogenous factors.
3 Of course, diplomatic “closeness” suggests other explanations for why the offended country doesn’t engage in retaliation.
Indeed, we contend that, in these sorts of situations, parties expend resources to make private information local: In the case of Toyota, for instance, exceptions are made on the basis of production conditions facing both the subcontractor and Toyota, and (as is well documented) Toyota expends significant resources—and expects its subcontractors to expend significant resources—so that each side of the transaction knows a great deal about the production conditions affecting the other party. Part of the duties of a department chair or dean—and we believe an important success factor for a chair or dean—is a willingness to understand the personal conditions that affect faculty members in the department or school. (And, of course, the same observation can be made about supervisors of work groups in general.) Now, of course, in both these cases, the information investment can be attributed in the first place as being driven by the desire to make “appropriate” decisions. But, we contend, a second reason is so that the second party can be a credible witness about whether the violation of some rule is legitimate.

To the best of our knowledge, the use of local information in this sort of situation has not been formally modeled. (But see the literature review following.) So the principal purpose of this paper is to provide a formal, stylized model of how local information might be employed. We verify the basic premise—that when information is local, a second party who has the information can be an effective and credible source in excusing or forgiving exceptions to rules based on (otherwise) private information—but go on from showing that this might be done to a more detailed examination of how. The paper runs as follows:

1. The basic model is set up in Section 2, with preliminary analysis conducted in Section 3. Leaving details for later, the basic structure is an assembly of players, where players sometimes have the opportunity to do a favor for another player. The players directly involved (the favor giver and favor receiver) know the levels of cost and benefit of the favor; the other players observe only that the opportunity for a favor takes place and whether the favor giver chooses to do the favor. Even if the pattern and frequency of favor opportunities is such that pairs of players cannot “trade” favors, community enforcement may be effective. High-cost, low-benefit favors may be inefficient, but as the community cannot directly tell which favors should not be done, if community enforcement is needed, the “cost” of having favors done is that inefficient favors may have to be done. This is our rules versus discretion conflict.

2. In our preliminary analysis of Section 3, the assumption that the favor receiver knows the cost and benefit of each favor is not used; while both the favor giver and receiver know these values, only the favor giver (she) acts. In Section 4, we consider allowing the favor receiver (he) to excuse the favor giver ex ante: Having seen the cost and benefit levels, he employs cheap talk to tell the community that he doesn’t expect this favor to be done. This can be used to enhance efficiency; in some cases efficient outcomes can be achieved. But in other cases, it is impossible to achieve full efficiency, even asymptotically.

For one thing, one can imagine that the leadership of the offended country simply wishes to be more lenient with allies than with countries with which it has a more adversarial relationship. And to the extent that “closeness” is correlated with a greater level of trade relationships, the countries may be engaging in bilateral reciprocity; see the theoretical developments following.
3. In Section 5, we imagine that instead of excusing the favor giver from doing the favor \textit{ex ante}, the favor receiver might publicly \textit{forgive} the favor giver for favors that are not done \textit{ex post}. This sort of construction works in the strong sense that efficiency can be achieved exactly for small enough \(r\) (if the basic parameters of the problem meet some simple conditions), even in cases where \textit{ex ante} excuses fail to achieve efficiency even asymptotically. This is because, once the favor has or has not been done, its particular benefits (and costs) are irrelevant to the continuation of the game.

4. Section 6 provides a comparison of \textit{ex ante} excuses and \textit{ex post} forgiveness, and Section 7 concludes with a discussion of extensions and variations on our basic model.\(^4\)

It should be noted at the outset that the notions of “rule” and “exception” are fluid; one can take the semantic position that we are merely investigating the replacement of a simple rule with a more nuanced rule that involves the exercise of some discretion in the application of the simple rule. We grant this semantic point, but we still believe the equilibria and conclusions we draw from them are economically interesting.

\textit{Related Literature}

We are unaware of papers that ask the same questions we are asking, although there are two strands of literature that are related and employ similar language.

A characterization of what we do here is that we are looking at the transmission of information from a small set of (two) players to a larger community. This in some ways echoes the substantial literature on random matching and the folk theorem. Much of this literature concerns the issues that arise when players who are matched do not automatically know what has been the history of play of the person against whom they are matched; in such circumstances, how can what happened in the past between pairs of players be adequately communicated to others, adequate (at least) to sustain folk-theorem like outcomes? This includes very early work by Rosenthal (1979) and, more representative of this literature, Kandori (1992) and Ellison (1994). The working paper by Wolitzky (2012) is of this type; his language is very close to ours. But despite some superficial similarities, there are substantial differences: this literature is concerned with issues of “hidden action” or moral hazard, where the past play of a player is not known by everyone. In our model, we assume that each player knows the (public-information) history of play by all the other individuals in the society; the issues that we address concern the ability to transmit (in credible form) information about exogenous variables that only some players hold.

The transmission of private information about payoff relevant variables via cheap talk is nearly as “old,” beginning with the seminal paper of Crawford and Sobel (1982). Sobel (2011) gives a summary of this literature. Especially relevant to our stylized model is the literature on “trading favors” (Mobius [2001]; Hopenhayn and Hauser [2004]) that deals with issues of “hidden information” and, similar to the model we develop, the question of parties doing favors for one another. But most of this literature

\(^4\) An Appendix provides the one very technical proof, and an On-line Appendix provides a number of bells and whistles that are mentioned but not fully discussed in the paper.
concerns one informed player transmitting information to a second player. A small literature, beginning with Krishna and Morgan (2004), concerns cheap-talk communication by multiple informed parties; see Sobel (2011, Section 5) for a more complete bibliography.

We began by invoking the literature on “rules versus discretion,” so connecting our model to this literature in macroeconomics may be worthwhile. In this literature (e.g., Kydland and Prescott, 1977), monetary authorities would like to announce a particular path of (contingent) action. If this announcement is believed, the general public will respond in a particular and desirable fashion. But once the public has responded in that fashion (anticipating that the authority will act as announced), the optimal course of action for the monetary authority changes; if the authority has discretion to act in its (new) best interests, it will not carry out the policy on which basis the general public originally acted. The general public, recognizing this, will not act initially on the supposition that the authority will keep to its announced policy, at least insofar as the authority retains discretion. So the monetary authority may choose instead to follow some rule, giving up its discretion. Barro and Gordon (1983) and Barro (1986) pick up this story, asking, in essence, what force exists to ensure that the authority will follow the “rule.” They invoke either an infinite-horizon or bounded-horizon-with-incomplete-information reputation construction: The authority, insofar as it is long-lived, keeps to the rule to preserve a reputation for doing so.

But for any reputation construction to work, where the point of the reputation is to give credibility to promised (or threatened) future actions and so to induce favorable current actions from others, it must be possible for the “general public”—the parties for whom this credibility is required—to verify ex post that the first party has behaved in accordance with its reputation. It is here that our model enters the story. We imagine that the information needed to verify ex post that the first party has conformed to a contingent rule is not widely held, so the contingent rule seemingly cannot be the basis of a viable reputation. (How) Can a second informed party provide for the general public the required verification?

In some institutional settings, the second informed party has as profession this task of verification. The second informed party, to the greatest extent possible, maintains its “independence” from the decisions of the first party, and so is trustworthy; think for instance of “independent auditors” (and the controversy in the audit industry of auditing firms that also have consulting arrangements with the firms they audit). In the model we explore, we look at a very different situation, namely where the second informed party, if anything, has the most at stake in the decision of the first party, which the second party is called upon to verify.

It may be helpful here to think of the case of Toyota and its subcontractors. Toyota wishes to maintain a reputation of being “fair” with its subcontractors. If it can credibly commit to behaving in this fashion in the future, it can induce subcontractors to make sunk-cost investments in the relationship. But subcontractors must worry that, having made those sunk-cost investments, Toyota will resort to a hold-up. Toyota, to be credible in its future actions, must put its reputation for being fair on the line with all its subcontractors simultaneously; together all subcontractors have sufficient bargaining power vis-à-vis Toyota to keep Toyota “in line.” And here our story enters: (How) Can Toyota and one of its
subcontractors, which share “local information” not available to the community of subcontractors, employ that information to their mutual advantage, without harming the reputation construction that is essential to the entire system?

2. A basic stylized model

We work with variations on the following stylized model.

There are $I$ players, indexed by $i = 1, \ldots, I$.

Time is continuous, indexed by $t \in [0, \infty)$.

Opportunities arise at random for players to do favors for one another, generated by independent Poisson processes. The rate of arrival of opportunities for $i$ to do a favor for $j$, hereafter called an $i$-for-$j$ favor, is denoted by $\lambda_{ij}$. Let $T^n_{ij}$ denote the arrival time of the $n$th opportunity for $i$ to do a favor for $j$. (Under the assumptions of independent Poisson arrivals, there is zero probability that at any time $t$ more than one favor opportunity takes place.)

Whenever an $i$-for-$j$ favor opportunity occurs, the cost to $i$ of doing this favor and its benefit to $j$ are determined randomly, independently of the (inter-)arrival time at which it applies. Think of there being, for each ordered pair $i$ and $j$, an i.i.d. sequence of pairs of real numbers $\{(x^n_{ij}, y^n_{ij}); n = 1, 2, \ldots\}$, each of these sequences fully independent of all other such sequences and independent of all the arrival times $T^n_{ij}$, such that at the time $T^n_{ij}$ of the $n$th $i$-for-$j$ favor opportunity, the cost of the favor to $i$ is $x^n_{ij}$ and the benefit $j$ receives if this favor is done is $y^n_{ij}$. We are not assuming that the values of $x^n_{ij}$ and $y^n_{ij}$ are independent of one another; we allow the distribution of each cost–benefit vector $(x_{ij}, y_{ij})$ to depend on the ordered pair $(i, j)$. We assume these distributions are such that:

- Each $x_{ij}$ and $y_{ij}$ is strictly positive with probability 1.
- The probability that $x_{ij} = y_{ij}$ is zero, for each $i$ and $j$.
- Each $x_{ij}$ and $y_{ij}$ has finite support.

The last of these assumptions is made to simplify some proofs; but an assumption that the supports of $x_{ij}$ and $y_{ij}$ are bounded is essential to several of our key results. Of course, the last assumption ensures that costs and benefits have finite expectation; so let

$$a_{ij} := E[y_{ij}] \quad \text{and} \quad b_{ij} := E[x_{ij}].$$

Also, let $m_{ij}$ be the highest value that $x_{ij}$ takes on with positive probability.

Available actions (for the time being) are simple: If, at time $t$, an $i$-for-$j$ favor opportunity occurs, then $i$ must decide whether to do the favor or not.

We assume that the only way one player can provide benefits to another is through these favors. In particular, we assume that utility transfers between the players are not possible. In Section 7, we briefly

---

5 Anticipating a bit, we use continuous time with Poisson arrivals of opportunities to do favors, to avoid a situation in which one player is called upon to do more than one favor at any single point in time.
reintroduce the possibility of monetary transfers, at which point we discuss this assumption in greater detail.

The key to our model is the distribution of information: We assume that every player knows when any $i$ has the opportunity to do a favor for some $j$ and, subsequently, whether $i$ does that favor or not. But, if $i$ has the opportunity to do a favor for $j$ at time $t$, the cost-benefit values $(x_{ij}, y_{ij})$ for this favor are common knowledge between $i$ and $j$ and unknown to all the other players.

The payoff for each player is the infinite-horizon discounted sum of the value of favors received less the cost of favors given, with an instantaneous interest rate of $r$. That is, a cost or benefit incurred at time $t$ is discounted by $e^{-rt}$. Each player seeks to maximize the expectation of her (infinite-horizon) payoff.

3. **Preliminary Analysis**

A variety of perfect equilibria can be constructed for the basic model. Or, put more correctly, we can describe a variety of strategy profiles and give conditions for them to be perfect equilibria.

**Autarky**

The simplest perfect equilibrium, which holds in all situations, is autarky: no one does any favors for anyone else; and all players get a payoff of 0.

**Bilateral reciprocity**

Suppose that some pair $i$ and $j$ agree to the following rule concerning their dealings with one another: $i$ does all favors she can for $j$, and $j$ reciprocates by doing all the favors he can for her, as long as each behaves in this fashion. If either fails to do a favor for the other, they never subsequently do favors for each other. This is just bilateral reciprocity between $i$ and $j$ as in the repeated prisoners’ dilemma, enforced by “grim punishment,” adapted to this setting. The expected flow of benefits to $i$ of this arrangement is $\lambda_{ij}a_{ij}$ while the expected flow of costs is $\lambda_{ij}b_{ij}$, so this is equilibrium behavior for $i$ as long as the immediate cost of doing any favor for $j$, which has upper bound $m_{ij}$, is less than the expected, discounted value of ongoing benefits less costs. That is, this constitutes (perfect) equilibrium behavior for $i$ and $j$ (in their bilateral relationship) as long as

$$m_{ij} \leq \frac{\lambda_{ij}a_{ij} - \lambda_{ij}b_{ij}}{r} \quad \text{and} \quad m_{ji} \leq \frac{\lambda_{ij}a_{ij} - \lambda_{ij}b_{ij}}{r}.$$  

(1)

Of course, this is but one possible bilateral arrangement between $i$ and $j$. For instance, suppose the relationship between $i$ and $j$ is symmetric: $\lambda_{ij} = \lambda_{ji}$, and the distributions of cost and benefit in $i$-for-$j$ favors and $j$-for-$i$ favors coincide. If, for some favors, the cost exceeds the benefit provided, $i$ and $j$ can agree between themselves not to do those favors.

---

6 By “perfect,” we mean “perfect Bayes,” although we won’t be precise about this. In fact, we are fairly confident that the equilibria we define are all sequential—we’ll say a bit more about this later—but we won’t verify that the consistency criterion holds.
Since payoffs in the overall game are additively separable across pairs, we can use this sort of bilateral reciprocity to construct perfect equilibria for the entire game: pairs $i$ and $j$ make arrangements between themselves that they can (bilaterally) enforce; if no such arrangement between some $i$ and $j$ is possible, they do no favors for one another.

**Two examples, and social enforcement**

Our interest is not in such aggregations of bilateral relationships, but instead in cases where bilateral relationships are largely inadequate to achieve efficient outcomes. Two specific examples illustrate what we have in mind.

**Example 1. The Circle, where no one serves the person who serves her.** For any $I \geq 3$, suppose that $\lambda_{ij} = 1$ if $j = i + 1$ and $\lambda_{ij} = 0$ if $j \neq i + 1$, where we interpret $I + 1$ as 1. In words, if we arrange the individuals in a circle, numbered $1, 2, \ldots, I$ as we go clockwise around the circle, each $i$ gets the opportunity to do favors for (only) her clockwise neighbor and receives favors (potentially) from (only) her anti-clockwise neighbor. To finish the example, suppose $r = 0.1$ and each $(x_{ij}, y_{ij})$ pair is degenerate at $(2,3)$, as in the previous example. The point is that, clearly, for no pair is bilateral reciprocity going to work; for pairs $i$ and $j$ where $j = i + 1$ or vice versa, one of the two inequalities in (1) holds, but the other fails; and for all other pairs $i$ and $j$, $\lambda_{ij} = \lambda_{ji} = 0$ (the two never interact), so in that sense, the inequalities are irrelevant.

**Example 2. Star-shaped relationships.** For, say, $I = 21$, suppose that $\lambda_{ij} = \lambda_{ji} = 0.1$ for $j \neq 1$, and $\lambda_{ij} = 0$ if neither $i$ nor $j$ is 1. In other words, everyone (other than 1) can give and receive favors from player 1, but no other pairs interact. Suppose $r = 0.1$, $(x_{ij}, y_{ij})$ has degenerate distribution at $(9,10)$, and $(x_{ij}, y_{ij})$ is degenerate at $(1,10)$, for each $j \neq 1$. For each $j \neq 1$, the expected present value of the flow of favors less costs (if 1 and $j$ carry out all favors) is $(0.1)(10 - 1)/(0.1) = 9$, more than enough so that $j$ has incentive to keep a bilateral all-favors arrangement with 1. But for 1 vis a vis $j$, the expected present value is $(0.1)(10 - 9)/(0.1) = 1$, which is insufficient to induce her to do favors of immediate cost 9, even if failing to do so means that some $j$ will never do favors for her in the future.

So, in both these examples, if we relied on bilateral reciprocity, no favors would be done. But suppose instead we construct equilibrium where players are punished by the whole community if they fail to do a favor. Keeping this as simple as possible, suppose that each $i$ is expected to (and, along the path of play, does) all favors that she has the opportunity to do, with the off-path threat that, if anyone fails to do a favor for anyone else, all players immediately move to autarky forever.\(^7\) In both examples, this is an

---

\(^7\) You might expect instead that, if $i$ fails to do a favor for $j$, then $i$ alone is punished, or that if punishment means autarky, the autarkic period is of finite duration. We construct equilibria of this character later in the paper; for now we are looking for the simplest possible all-favors equilibrium.
equilibrium: In place of inequality (1), we require that for player \( i \),

\[
\max_{j \neq i} m_{ij} \leq \frac{\sum_{j \neq i} [\lambda_{ji} b_{ji} - \lambda_{ij} a_{ij}]}{p} \quad \text{for each } i.
\]

If these inequalities hold, each player (along the path of play) prefers to incur the immediate cost of any favor she might be called upon to do, if this preserves the all-favors arrangement and the alternative is autarky. (Of course, the off-path threat of autarky is perfect Bayes.) These inequalities hold in the two examples: In the circle example, \( i \) does favors for \( i + 1 \) so that \( i - 1 \) will continue to do favors for her; in the star-shaped arrangement, player 1 is willing to do favors costing her 9, to preserve the expected net flow of value she receives the twenty relationships she has with the various \( j = 2, \ldots, 21 \).

But all-favors may be inefficient

Suppose that, in the context of Example 1 (the circle), cost–benefit vectors for each pair \((i, j)\) (where \( j = i + 1 \)) have the following probability distribution: With probability 0.8, \((x_{ij}, y_{ij}) = (2, 3)\); with probability 0.1, \((x_{ij}, y_{ij}) = (1, 4)\); and with probability 0.1, \((x_{ij}, y_{ij}) = (5, 4)\). The average values are 2.2 for \( x_{ij} \) and 3.2 for \( y_{ij} \), hence in the circle example, all-favors with the threat of autarky is a perfect equilibrium with expected payoffs of 10 for all players. (Bilateral arrangements always fail in the circle example.) But it is inefficient: The third type of favor, with cost–benefit vector \((5, 4)\), would be better left undone; if we could find a way to do only favors of the first two types, the expected payoff for each party (with the rest of the data taken from the circle example) would be 11.

For each \( i \) and \( j \), let \( S_{ij} \) denote the set of all subsets of the support of \((x_{ij}, y_{ij})\), with \( S_{ij} \) a typical element of \( S_{ij} \). And let

\[
\Sigma := \prod_{(i,j):\lambda_{ij}>0} S_{ij}, \quad \text{with typical element } S = (S_{ij})_{(i,j):\lambda_{ij}>0}.
\]

In words, an \( S \) is a selection of all possible favor “types,” where one favor type is distinguished from another by who are the giver and receiver, and what are the cost and benefit. Our objective is see whether, with community enforcement being the “hammer” that keeps players in line, a particular selection \( S \) can be implemented as a perfect equilibrium among the players, in the sense that, along the path of play, the favors that are done are precisely those that are in the selection \( S \).

One selection \( S \) is of particular interest: Let

\[
S_{ij}^U = \{(x, y) \in \text{the support of } (x_{ij}, y_{ij}) : x < y\} \quad \text{and} \quad S^U = (S_{ij}^U)_{(i,j):\lambda_{ij}>0}.
\]

That is, \( S^U \) is the selection in which the favors that are (meant to be) done are those whose cost is less than the benefit they provide; of course, the outcome that results from this selection will maximize the sum of
expected payoffs of all the players; the superscript $U$ is for *utilitarian*, and we use the term *u-efficient* to refer to both the selection $S^U$ and the outcome (in terms of payoffs) it engenders.

Insofar as we rely on social enforcement, it is not simple to arrange that the players do precisely the favors in a selection $S$ that is less than “all favors.” When an *i*-for-*j* favor opportunity occurs, *i* and *j* will know whether its cost and benefit qualify it as part of the selection. But other players will not know this. If the favor is not in the selection, *i* should not be “punished” for failing to do it. But, then, how will the wider community know whether to punish *i* for an undone favor? If *i* goes unpunished for not doing the favor, then *i* has no incentive to do any favor, so some punishment (or, alternatively, some reward for doing the favor) is required.

This is the rules-versus-discretion dilemma within our model. “Do all favors” allows for social enforcement, because everyone can verify ex post whether this rule is being followed. But this rule precludes the parties immediately concerned from exercising discretion and employing information they possess (privately or locally) to enhance efficiency or otherwise make a desirable selection.

In fact, the previous paragraph overstates matters. Equilibria can be created in which the favor giver unilaterally exercises discretion and forgoes doing some favors, a choice for which the favor giver is subsequently punished. Given the nature of the punishment, the favor giver may, for some very costly favors, prefer to be punished. But note that in such a situation, it is the cost of the favor that drives the favor giver’s decision; if a selection, on efficiency grounds, includes some favors with greater cost than some not in the selection, unilateral discretion cannot work.

Analysis of unilaterally exercised discretion takes us too much out of our way; we want to involve the favor receiver, who also knows the costs and benefits of a given favor, in the exercise of discretion. So we will leave discussion and analysis of unilateral discretion for the concluding remarks and the on-line appendix.

*Notation for a fixed selection $S$*

Some notation is helpful in dealing with a fixed selection $S$. For a given $S$, and for pairs *i* and *j* such that $\lambda_{ij} > 0$, let

\[ A_{ij}(S) := \mathbb{E}[y 1_{(x,y) \in S_{ij}}], \quad B_{ij}(S) := \mathbb{E}[x 1_{(x,y) \in S_{ij}}], \quad \text{and} \quad M_{ij}(S) := \max\{x : (x,y) \in S_{ij}\}. \]

That is, $A_{ij}(S)$ is the expected benefit *j* derives from a favor opportunity where *i* is the favor giver and if only favors in the (sub)selection $S_{ij}$ are performed, $B_{ij}(S)$ is the expected cost to *i* of a *i*-to-*j* favor opportunity, if only favors in $S_{ij}$ are done, and $M_{ij}(S)$ is the most costly favor *i* is called upon to do for *j*, under the selection $S$. (If *i* is called upon to do no favors for *j* in $S$, set $M_{ij}(S) = 0$.) And define

\[ A_i(S) := \sum_{j \notin i} \lambda_{ji} A_{ji}(S) \quad \text{and} \quad B_i(S) := \sum_{j \notin i} \lambda_{ij} B_{ij}(S). \]
That is, $A_i(S)$ is the expected flow rate of benefits accruing to $i$ if favors in the selection $S$ are done for her, and $B_i(S)$ is the expected flow rate of costs she accrues, if she does all the favors she is meant to do in the selection $S$.

**Social strategy profiles and perfect social equilibria**

Before moving to analysis, a final definition is required.

**Definition.** A strategy profile is **social** if it consists of (behavior) strategies for each player in which the actions taken by each player at any point in time depend only on information that is either part of the common-knowledge history of play or is “current.” More precisely, we will look at various game forms in which actions only take place at the moment that a favor opportunity occurs and where, moreover, only the $i$, $j$ pair relevant to the particular favor act. In formulating their actions at the moment of a particular $i$-for-$j$ favor opportunity, this $i$ and $j$ can condition only on the public or common-knowledge history of the game and the cost and benefit levels of the current favor opportunity. **We use the phrase perfect social equilibrium as shorthand for a perfect equilibrium involving a social strategy profile.**

For the remainder of the paper, we will investigate only perfect social equilibria. There can be no doubt that this is serious restriction: For instance, strategy profiles that employ bilateral reciprocity will not in general be social; bilateral reciprocity in general involves strategies where, if $i$ fails to do a favor for $j$ that, based on its benefit and cost, $i$ was meant to do, $j$ punishes $i$ by withholding later favors. As long as $i$ and $j$ are not alone, and absent any credible broadcast by $i$ and/or $j$ about the cost–benefit vector, the particular values for the cost and benefit do not become public information. So why restrict the strategy profiles we are willing to consider in this fashion? Our interest is in cases where enforcement requires social action. It may be possible to construct equilibrium strategies in which the continuation play (after some action) is not common knowledge to all the players and yet that involves some level of social enforcement. But such equilibria are beyond our abilities to study. Put more positively, we will show how, by involving the favor receiver in cases where a favor is undone, we can implement some selections $S$ with a perfect social equilibrium. To the extent that having social strategies is a virtue, this sort of result is strengthened by the restriction.

A further excuse we can offer for this restriction is that it has a long history in the literature; it simplifies the analysis of specific strategies because of the following result (the proof of which is obvious).

**Lemma.** In any social equilibrium, the continuation payoffs for all players following the immediate interaction of a pair engaged in a favor opportunity depend only on the public (or common knowledge) history of the game.

---

8 Ghosh and Ray (1996) use the same name for a very different object.

9 Please note that a “perfect equilibrium involving a social strategy profile” is a social strategy profile in which each player, using a social strategy, is playing an unrestricted best response against the strategies of her opponents. In testing whether we have an equilibrium, we do not require that deviations are social.

10 Once we restrict attention to social strategy profiles, this lemma allows us to be more formal concerning what perfection
4. Ex Ante Excuses

We come to the main question of this paper: If the information needed to achieve efficiency is local, meaning held by both the potential favor giver and receiver, (how) can the potential favor receiver be enlisted to help implement (as a perfect social equilibrium) a selection that is smaller than all-favors?

We do not change the rule that says that the potential favor giver must decide unilaterally whether to grant the favor. We do not change any payoffs, contingent on the choices favor givers make in that regard. But we will allow the favor receiver to issue public, cheap-talk statements, intended to guide the behavior of others.

Of course, cheap talk can take many forms. We could imagine the potential favor giver and receiver both engaging in cheap talk.\footnote{In the context of this game, where both know precisely the cost and benefit of a given favor, and where there is a "punishment" that is both dire and, on its own, an equilibrium, simultaneous cheap talk can easily used to implement a selection $S'$, assuming that Condition A in Proposition 1 holds for $S'$. Prior to the moment when the favor must be done or not, but after both the favor giver and receiver have seen the pair $(x, y)$, they simultaneously announce "It should be done" or "It should not." If their messages coincide, what they announce happens. If their messages disagree, autarky immediately ensues, forever. If there is any deviation after a common message (say, they announce "It should be done" and then it is not), autarky immediate ensues. There are well-known strategic issues arising from this sort of equilibrium; e.g., each pair of favor giver and receiver, if they can address one another privately prior to the announcement, can threaten the other. And there are more subtle issues to resolve if their information is not identical; one then looks for whether, with a richer language, approximate efficiency can be achieved. We will not pursue this further; in this paper, the only cheap talk is the favor receiver.}

Even if we restrict to cheap talk by the receiver only, we could imagine him talking before the favor decision is made and then again after. And the language employed could be quite large.

In this paper, we look at the following simple situations. The only party that engages in cheap talk is the favor receiver. In this section, he speaks before the favor decision is made and in places is limited to saying either “The favor giver is excused, this time” or “Not excused” (or, what is equivalent, he says nothing). Next section, he speaks after the favor decision and, at least for some results, he may speak only if the favor is not done, at which point he says either “The favor giver is forgiven, this time,” or not. Our justification is that these possibilities already raise a number of subtle and interesting issues to be explored before (perhaps) going on to more complex patterns and language of cheap talk.

We reiterate a point made in the Introduction: Cheap-talk declarations, especially ex post, by otherwise disinterested parties are a common phenomenon. (In transaction-cost economics, the term \textit{trilateral governance} is sometimes used to describe such situations.) Auditors, for instance, are meant to verify (after the fact) that financial reports issued by a company are accurate. As a (theoretically) disinterested party, auditors have no incentive to do anything other than to tell the truth.\footnote{Of course, were they completely disinterested, they would have no incentive to say anything in particular; truth-telling is but one equilibrium. And insofar as effort is required to get at the truth, they have no incentive to expend effort, until one enriches requires. In a social equilibrium, we worry about behavior involving a pair $i$ and $j$ at an instant at which an $i$-for-$j$ favor opportunity arises—we never will consider situations in which anyone other than $i$ and $j$ act at that instant—and about how the game progresses after the dust of this sort of interaction clears. In a social equilibrium, actions subsequent to a particular favor interaction are based on a history of the game that is common-knowledge to all players; this isn’t quite subgame perfection in the formal sense, but it is effectively the same. And within a particular interaction, we assume that $i$ and $j$ are on a common-knowledge basis, at least insofar as their actions are concerned: the past affects current actions only through the common-knowledge elements of their history, while the current information of costs and benefits of the immediate favor are, by assumption, known to both.

In our story, however, it
will be up to the other party most directly affected by the decision—the favor receiver—who must make the cheap-talk declaration. We find this interesting on grounds that this is the party most likely to share in the information required for efficient arrangements.

With ex ante cheap talk, the question Can selection $S$ can be implemented (for a given $r$) in a perfect social equilibrium? has a simple answer.

**Proposition 1.** If favor receivers are able to issue cheap-talk declarations (only) before the decision whether to do the favor, a selection $S$ can be implemented for a given $r > 0$ if and only if, for all $i$ and $j$ such that $\lambda_{ij} > 0$

$$M_{ij}(S) \leq \frac{A_i(S) - B_i(S)}{r} \quad \text{and}$$

$$(x, y) \notin S_{ij} \text{ implies } x \geq M_{ij}. \quad (B)$$

Condition A is fairly obvious (or will be, once you see the proof); this is the incentive compatibility constraint along the path of play. The real meat of the proposition is Condition B, which says that the selection $S$ must have a “cut off” structure in terms of costs: Favors that are to be done must be (weakly) less costly than are favors that are not done, for each ordered pair of players.

**Proof.** Since all favor types have a strictly positive cost, condition A holding implies that either $A_i(S) > B_i(S)$, or $i$ is called upon to do no favors in the selection $S$. In the latter case, implementation of $i$’s role is trivial. So we will assume that $A_i(S) > B_i(S)$ for all $i$, leaving the case where this strict inequality holds for only a subset of the $i$’s to the reader.

We first show that if $S$ satisfies A and B, we can construct a simple cheap-talk regime that implements the selection $S$: For each $i$ and $j$ let

$$\phi_{ij} := \frac{M_{ij}(S)r}{A_i(S) - B_i(S)}.$$ 

Note that, per the assumption of the proposition, $\phi_{ij} \geq 0$. Specify the following strategies. When an $i$-for-$j$ favor opportunity arises with cost–benefit vector $(x, y)$:

1. If $(x, y) \in S_{ij}$, $j$ does not excuse $i$ and $i$ does the favor.
2. On the other hand, if \((x, y) \notin S_{ij}\), then \(j\) excuses \(i\), \(i\) does not do the favor, and all await the next favor opportunity to come along.

3. If \((x, y) \in S_{ij}\), and \(j\) (by mistake) forgives \(i\), \(i\) doesn’t do the favor.

4. Regardless of the values of \(x\) and \(y\), if \(j\) does not forgive \(i\) and \(i\) does not do the favor, then a publicly observable randomization is conducted where, with probability \(1 - \phi_{ij}\), all players “ignore” what just happened, while with probability \(\phi_{ij}\), everyone moves to autarky.\(^{13}\)

These strategies are clearly social, and they implement the selection \(S\) along the path of play. We must check whether they are perfect equilibrium strategies. If ever play moves to autarky, continued play is clearly in equilibrium. So suppose an \(i\)-for-\(j\) favor opportunity arises with cost–benefit pair \((x, y) \in S\). If \(j\) does not excuse \(i\), \(i\) is supposed to do the favor, so to get the benefit of this favor, \(j\) will not excuse \(i\). And if \(j\) does not excuse \(i\), \(i\) can do the favor, with an immediate payoff plus continuation value of

\[
\frac{A_i(S) - B_i(S)}{r} - x,
\]

or she can fail to do the favor, with a continuation value of

\[
(1 - \phi_{ij}) \left[ \frac{A_i(S) - B_i(S)}{r} \right] + \phi_{ij} \cdot 0 = \frac{A_i(S) - B_i(S)}{r} - \phi_{ij} \left[ \frac{A_i(S) - B_i(S)}{r} \right]
\]

\[= \frac{A_i(S) - B_i(S)}{r} - M_{ij}.\]

Since \(x < y\), \(x \leq M_{ij}\) by definition, and \(i\) is content to do the favor.

On the other hand, suppose that \((x, y) \notin S\). The favor receiver \(j\) is supposed to excuse \(i\). If he does, the favor will not be received, but play will continue along the equilibrium path. And if he doesn’t excuse \(i\), \(i\) will still not do the favor, running the risk of triggering autarky. So excusing \(i\) is clearly a best response for \(j\), given \(i\)’s strategy. As for \(i\), she is supposed not to do the favor whether excused or not. If she is excused, she can forego doing the favor with no adverse consequences (nor does she gain anything by doing the favor), so she won’t do it. But what if \(j\) fails to excuse her? The same comparison of payoffs as in the case where \((x, y) \in S\) and \(i\) is not forgiven apply, but now, since \((x, y) \notin S\) implies that \(x \geq M_{ij}(S)\) (Condition B),

\[
\left[ \frac{A_i(S) - B_i(S)}{r} \right] - M_{ij} \geq \left[ \frac{A_i(S) - B_i(S)}{r} \right] - x,
\]

\(^{13}\) It will be clear from the proof that this is an equilibrium that, alternatively, we could in these circumstances have all players move to autarky for a length of time that inflicts the same expected costs on \(i\) as does this random descent into autarky. A more interesting issue is that this “punishment” is inflicted not only on \(i\), but as well as on \(j\) and on everyone else. We’ll discuss this point after the proof.
where the left-hand side is $i$’s expected payoff if she doesn’t do the favor and the right-side side is her payoff if she does. She (weakly) prefers not to do the favor, showing that our purported equilibrium is indeed a (perfect) equilibrium.

Conversely, suppose that Condition A does not hold. Then there is some ordered pair $(i, j)$ and an $i$-for-$j$ favor $(x, y) \in S_{ij}$ such that $x > [A_i(S) - B_i(S)]/r$. But, then, suppose an $i$-for-$j$ favor happens with cost-benefit vector $(x, y)$: $i$’s continuation payoff is $[A_i(S) - B_i(S)]/r$ (if we are implementing $S$), so doing this favor makes $i$’s overall payoff strictly less than 0. By refusing to do all favors now and in the future, $i$ is assured of a payoff of at least 0. Hence doing this favor is not a best response by $i$.

Finally, suppose that Condition B does not hold. That is, for some ordered pair $(i, j)$ with $\lambda_{ij} > 0$, there is an $i$-for-$j$ favor type $(x, y)$ with $x < M_{ij}(S)$. Suppose $S$ could be implemented in a perfect social equilibrium with ex ante cheap talk. Let $(x', y') \in S$ be an $i$-for-$j$ favor type where $x' = M_{ij}(S)$. When an $i$-for-$j$ favor opportunity arises (which happens with probability 1 infinitely often), it could be of type $(x, y)$ and it could be of type $(x', y')$. Since this perfect social equilibrium implements $S$, there is some cheap-talk statement that $j$ can make, call it $M$, that causes $i$ (in equilibrium) to do the favor with probability 1, if the favor is of type $(x', y')$. Note that, after this happens, $i$ has the continuation value $V^* = [A_i(S) - B_i(S)]/r$ and $j$ has the continuation value $U^* := [A_j(S) - B_j(S)]/r$, since this equilibrium implements $S$. Let $V$ be $i$’s continuation value if she does not do the favor; since she is willing to do the favor, it must be that $V^* - x' \geq V$.

Now consider if the $i$-for-$j$ favor is of type $(x, y)$. Since the path of play implements $S$, there is something $j$ says that allows $i$ not to do the favor. After saying this (and not getting the favor), $j$’s continuation value is, once again, $U^*$.

So suppose that the favor is of type $(x, y)$, and $j$ issues the cheap-talk declaration $M$. Player $i$ must choose between doing the favor, for a payoff of $V^* - x$, or not doing it, for a payoff of $V$. (Here the assumption that the strategies are social comes into play: Continuation payoffs can only depend on what $j$ says and whether or not $i$ does the favor.) Since $x < x'$ and $V^* - x' \geq V$, it follows that $V^* - x > V$, so $i$ will surely do the favor (because the equilibrium is perfect). And, then, $j$ gets the value $y > 0$ of the favor, and same continuation value $U^*$ that he would have had, had he followed the prescribed equilibrium. Therefore, this was not an equilibrium.

Note that in the equilibrium we construct (if A and B hold), everyone suffers if $i$ fails to do a favor for which she is not excused. (Of course, this is an out-of-equilibrium occurrence; it never happens along the path of equilibrium play.) It is unnecessary for the “innocent bystanders”—everyone except for $i$ and $j$—to suffer, and depending on the detailed structure in a particular example, it may be possible to arrange matters so that only $i$ and $j$ suffer.\footnote{But care must be taken here. In, for instance, the circle example, punishing any player in a way that removes her incentives to do favors destroys any chance of any favors being done.} But to maintain the equilibrium, $i$’s punishment cannot be designed in a way that benefits $j$ and, in fact, we suggest that $j$ ought to be punished as well as $i$:}

\footnote{But care must be taken here. In, for instance, the circle example, punishing any player in a way that removes her incentives to do favors destroys any chance of any favors being done.}
Player $i$’s punishment is set so that she will do all favors in $S_{ij}$ and, whether excused or not, will do no favors not in $S_{ij}$. It is then $j$’s responsibility to excuse her, if the favor type is not in $S_{ij}$. If $j$ does not excuse $i$ and $i$ does not do the favor, *either* $i$ or $j$ has failed to follow the equilibrium prescription, and third parties cannot tell who it is. So it is somewhat natural to punish both $i$ and $j$. Put another way, if $j$ suffers in these circumstances, $j$ has a strict incentive to forgive $i$ when an $i$-for-$j$ favor type is not in $S_{ij}$. To the extent that equilibria with strict best responses are more robust, punishing $j$ in these circumstances is a good idea.

A simple corollary to Proposition 1 is:

**Corollary.** Suppose that, for a given selection $S$, $A_i(S) > B_i(S)$ for all $i$ and Condition B holds. Then for some $r(S) > 0$, $S$ can be implemented in a perfect social equilibrium for all $r \leq r(S)$.

Of course, if the selection $S$ has $A_i(S) < B_i(S)$ for any $i$, then we cannot come “close” to implementing $S$, even as $r$ approaches zero. Leaving aside cases where $A_i(S) = B_i(S)$ for some $i$, one question remains: Suppose that the selection $S$ satisfies $A_i(S) > B_i(S)$ for all $i$; the selection of favors gives each player a strictly positive payoff, but Condition B fails. In these circumstances, we can’t implement $S$ as a perfect social equilibrium for any $r > 0$ if Condition B fails, but perhaps, in the spirit of the Folk Theorem, we can come asymptotically close, for $r$ approaching zero.\(^{15}\) We conjecture that this is true if the language that the favor receiver can employ in his cheap talk is sufficient rich. But, at least for one selection of interest (the u-efficient selection), we can prove the following anti-folk theorem:

**Proposition 2.** Suppose we restrict potential favor receivers to cheap talk only before the favor decision is made and, moreover, the favor receiver is limited to his choice of two (cheap-talk) messages before $i$ acts, such as that the favor giver is excused or not. And suppose that, for some pair $i$ and $j$ with $\lambda_{ij} > 0$, there are four cost-benefit pairs, $(x(k), y(k))$ for some $i$ and $j$ with

$$y(1) > x(1) > x(2) > y(2) > y(3) > x(3) > x(4) > y(4).$$

Then the u-efficient selection $S^U$ cannot be “implemented asymptotically”: There exists a strictly positive constant $K$ such that, if $(v_1, \ldots, v_I)$ is a vector of expected payoffs to the players in any perfect social equilibrium for interest rate $r$,

$$r \sum_i v_i \leq \sum_i [(A_i(S^U) - B_i(S^U)] - K.$$

\(^{15}\) To be clear, when we speak of implementing $S$, we have a time-homogeneous path of play: Every time a favor opportunity arises along the path of play, it is done if and only if it is in $S$. Here we are asking whether, for any sequence $\{r_n\}$ with limit 0, we can find a sequence of time-inhomogeneous social strategy arrays, where the $n$th of these is a perfect equilibrium for $r_n$, that (as $n \to \infty$) give in the limit the same payoffs as would implementation of the selection $S$. 

16
That is, the sum of normalized equilibrium payoffs in any perfect social equilibrium is uniformly bounded away from the sum of normalized payoffs from the selection $S^U$.

The proof is fairly technical, but it contains novel techniques for an “anti-folk theorem;” so we include it in the appendix, along with further discussion of why we believe that, with a rich cheap-talk vocabulary, “asymptotic implementation” is possible.

5. Ex Post Forgiveness

To try to implement selections that do not satisfy Condition B—and because it is of independent interest—we turn now to cheap talk by the prospective favor receiver after the favor giver has made her choice whether to do the favor. We have in mind that, if a favor goes undone, the (now disappointed) favor receiver can either publicly forgive the favor giver or not.

Ex post cheap talk has one major advantage over ex ante cheap talk. Cheap talk undertaken before the decision whether to do the favor must take into account the cost and benefit levels of the current favor; the favor could still be done, so those values are still germane. But with ex post cheap talk, the “current” favor is now in the past and so is irrelevant to continuation values.

On the other hand, ex post forgiveness of the sort we are seeking is delicate in at least one respect. Reason as follows: In a perfect social equilibrium, favor receiver $j$ will have two continuation values at a given point in time if, at that time, $i$ has the opportunity to do a favor for $j$ and fails to do so: $j$ has a continuation value if he forgives her, and he has a continuation value if he does not. Since the decision not to do the favor at this point is fait accompli, $j$ will do whichever of these (forgive or not) has the higher continuation value. Only if the continuation values are identical will he discriminate between the two responses.

But if the selection $S$ requires $i$ to do some favors for $j$ but not others, and $j$ is meant to forgive $i$ when she doesn’t do a favor that is not in $S_{ij}$, we need $j$ to employ both responses. If $i$ fails to do a favor that she is not supposed to do, we need $j$ to forgive her, without adverse consequences to her. So $j$’s continuation value of forgiveness must be at least as large as for not forgiving $i$. But if it is strictly larger, then $i$ can safely forbear from doing any favors for $j$, knowing that $j$ (in a perfect social equilibrium) will forgive her.

This suggests the following form for equilibrium strategy profiles, in implementing a selection $S$ in which some $i$-for-$j$ favors are done and others are not. When an $i$-for-$j$ favor opportunity arises with cost–benefit vector $(x, y)$, if $(x, y) \in S_{ij}$, then $i$ does the favor. If $(x, y) \not\in S_{ij}$, $i$ does not do the favor, and $j$ forgives her publicly. And, to provide $i$ with the incentive to do favors for which $(x, y) \in S_{ij}$, if $i$ fails to do such a favor, $j$ does not forgive her, and she is punished for this in a way that gives her the incentive to do all such favors. That is, $i$’s continuation value if she fails to do a favor and is not forgiven is less than her continuation value after she does a favor by an amount greater than $M_{ij}(S)$. Finally, $j$ is indifferent ex post between excusing an omitted favor and not, so he will always tell the “truth,” forgiving those omitted favors for which $(x, y) \not\in S_{ij}$ but not those that are $\in S_{ij}$.
How do we make \( j \) indifferent ex post? Suppose the punishment that is devised for \( i \) (if she fails to do a favor and is not forgiven) is such that \( j \) weakly benefits from it. If this is so, then if \( i \) fails to do a favor and is not forgiven, we can conduct a publicly observable randomization between punishing \( i \) and moving everyone to autarky; as long as autarky is worse for \( j \) than not punishing \( i \) (which it must be, for individual rationality to hold), a randomizing probability can be selected that exactly balances the continuation values for \( j \) if she forgives \( i \) and the expected value if she does not and either \( j \) is punished or autarky descends.

The key, then, is in devising punishment regimes. There are a number of ways this might be done; all the ones we have been able to construct have fairly complex off-path dynamics, but one is distinguished in that it gives the following remarkable result.

**Proposition 3.** Selection \( S \) can be implemented as a perfect social equilibrium with ex post cheap talk if and only if,

\[
\max_{j \neq i} M_{ij}(S) \leq \frac{A_i(S) - B_i(S)}{r}, \text{ for all } i. \tag{A}
\]

Moreover, the cheap-talk vocabulary required along the path of (equilibrium) play involves favor receivers saying (only) “\( i \) is forgiven, this time” or “no forgiveness,” if \( i \) does not do a favor, and the second of these two is never used along the path of play. (Favor receivers are also called upon to issue cheap-talk declarations in off-the-path situations, and as with along-the-path declarations, a vocabulary of two possible messages suffices.)

**Proof.** To take the negative half of the result first, if \( S \) is implemented in a perfect social equilibrium, then the payoff to player \( i \) is \((A_i(S) - B_i(S))/r\). Called upon to do a favor from \( S \), \( i \)’s continuation payoff must exceed the cost of this favor. So Condition A is certainly necessary for the implementation of \( S \).

For the positive half, we construct a perfect social equilibrium that implements \( S \). To motivate this construction, note that Condition A gives us no “slack”: Suppose Condition A holds with equality for player \( i \). Let \( j \) be the corresponding player for whom \( M_{ij}(S) \) attains the maximum, and let \((x, y)\) be the \( i \)-for-\( j \) favor type that has \( x = M_{ij}(S) = (A_i(S) - B_i(S))/r \). Since we are implementing \( S \), if player \( i \) does this favor, as she is meant to do, her continuation payoff will be \((A_i(S) - B_i(S))/r \). So if she does not do this favor and is punished, her continuation payoff must be 0. (It can’t be less, because 0 is her maxmin. And it can’t be more, or she won’t do this favor.) What we will do, then, is to design punishments for players in which their continuation values (if they are in the midst of punishment) is always 0.

We do this with an six-part description of the strategies the players employ:

1. Each player at any point in the game will be in one of two states, G (for “grace”) or P (for “purgatory”).
   All players start in state G.
2. If an $i$-for-$j$ favor arrives with cost-benefit vector $(x, y)$ when $j$ is in state $G$, then $i$ does the favor if $(x, y) \in \mathcal{S}_{ij}$ but does not do the favor if $(x, y) \notin \mathcal{S}_{ij}$.

3. If an $i$-for-$j$ favor arrives, when $i$ and $j$ are both in state $G$, and if $i$ does the favor, the game simply continues to the next favor opportunity, with no change in anyone’s status.

4. If an $i$-for-$j$ favor arrives when $i$ and $j$ are both in state $G$, and if $i$ does not do the favor, then $j$ issues a cheap-talk declaration: He forgives $i$ if the favor is not in $\mathcal{S}_{ij}$, and he does not forgive her is the favor is in $\mathcal{S}_{ij}$. If $j$ forgives $i$, the game continues to the next favor opportunity, with no change in anyone’s status. If $j$ does not forgive $i$, then a publicly observable randomization takes place with one of two outcomes: either $i$ is sent to state $P$; or everyone moves to autarky, forever.

5. If an $i$-for-$j$ favors arrives when $j$ is in state $P$, $i$ does not do the favor, regardless of the type of favor. Whether $i$ does the favor (by mistake) or not, play continues to the next favor opportunity with no change in status for any player.

So, a part of $i$’s punishment (when she has been sent to $P$) is that no one does any favors for her, unless and until she returns to state $G$. But we still want her to do favors for other players who are in $G$, at least for those favors that are called for under $\mathcal{S}$. The final three parts all concern how this happens; that is, what takes place if $i$ is in $P$, $j$ is in $G$, and an $i$-for-$j$ favor opportunity arises:

6. If an $i$-for-$j$ favor arrives when $i$ is in state $P$ and $j$ is in state $G$, $i$ does the favor if it is in $\mathcal{S}_{ij}$ and does not do it if it is not in $\mathcal{S}_{ij}$.

   a. If $i$ does not do the favor, play continues to the next favor opportunity, with no change in the status of any player.

   b. If the favor is of type $(x, y) \notin \mathcal{S}_{ij}$ and $i$ does the favor, $j$ issues a cheap-talk declaration that “$i$ is not restored to grace.” A publicly observable randomization is then conducted: Either no change in the status of any player occurs, or player $j$ is sent to state $P$. The probabilities for this randomization will be specified in a bit.

   c. If the favor is of type $(x, y) \in \mathcal{S}_{ij}$ and $i$ does the favor, $j$ issues one of two cheap-talk declarations: Either $j$ says “$i$ is restored to grace,” in which case $i$ is moved back to state $G$, and the game proceeds, or $j$ says “$i$ is not restored to grace,” in which case a publicly observable randomization takes place, with one of two consequences: No one changes status (and we proceed to the next favor opportunity), or $j$ joins $i$ in state $P$. The exact probabilities for $j$’s choice of which cheap-talk proclamation to make and for the publicly observable randomization will be specified in a bit.

We assert that these strategies are social. Specifically, after the dust settles on each favor opportunity, the publicly available information is sufficient to tell the state ($G$ or $P$) of each player, which then is sufficient to tell what happens at the next favor opportunity.

If players follow these strategies, starting from when all players are in state $G$, each player $i$ has an expected payoff of $(\mathcal{A}_i(S) - \mathcal{B}_i(S))/r$, which is also her continuation payoff if all are in state $G,
following any favor opportunity. If $i$ is in state $G$ and some others are in state $P$, $i$’s continuation payoff is (weakly) larger than this, because $i$ may be able to avoid doing some favors for those players in state $P$, at least for a while.\textsuperscript{16} But note that, as long as a player is in state $G$ and remains there, all favors specified by $S$ that are to be done for him, will be done for him.

We need notation for the continuation values of each player $i$, as a function of the set of players (other than $i$) who are in state $P$ and $i$’s state. Let $\mathcal{I}$ denote a generic subset of the set of players less $\{i\}$, including the null set, and write $v^G_i(\mathcal{I})$ for the continuation value for player $i$, if $i$ is in state $G$ and the players currently in $P$ are given by $\mathcal{I}$. And let $v^P_i(\mathcal{I})$ be $i$’s continuation value, if she is in state $P$ together with the members of $\mathcal{I}$.

We can now specify the probability in the private randomization that $j$ is meant to conduct in part 6c. Recall that the situation is that $i$ is in state $P$ and $j$ is in state $G$, and the favor is of some type $(x, y) \in S_{ij}$. Let $\mathcal{I}$ denote the set of players other than $i$ who are currently in state $P$. Then $j$’s private randomizing probabilities are $x/v^G_i(\mathcal{I})$ for “$i$ is restored to grace,” and the complementary probability for “$i$ is not restored to grace.” Note that, since we know that $v^G_i(\mathcal{I}) \geq (A_i(S) - B_i(S))/r$ and that $x \leq M_{ij}(S) \leq \max_{j' \neq i} M_{ij'}(S)$, Condition A tells us that this probability is between zero and one.

Will $i$ do the favor (if she is in state $P$, $i$ is $G$, and the favor type $(x, y) \in S$)? If she doesn’t do it, she remains in $P$, with continuation value $v^P_i(\mathcal{I})$. If she does, her expected payoff is

\[-x + \frac{x}{v^G_i(\mathcal{I})} \times v^G_i(\mathcal{I}) + \left[1 - \frac{x}{v^G_i(\mathcal{I})}\right] \times \left[\phi v^P_i(\mathcal{I}) + (1 - \phi) \times v^P_i(\mathcal{I} \cup \{j\})\right],\]

where $\phi$ is the public-randomization probability still to be specified for part 6c. To explain, in the display we have the cost of the favor, plus the probability that she is restored by $j$ to $G$, times her continuation value there, times the probability that $j$ does not restore her to $G$, times the quantity: her expected continuation value, depending on whether or not the public randomization causes $j$ to join her and the members of $\mathcal{I}$ in state $P$.

Suppose $v^P_i(\mathcal{I}) = v^P_i(\mathcal{I} \cup \{j\}) = 0$. Then the displayed expression is 0, which is $v^G_i(\mathcal{I})$, and $i$ is indifferent between doing the favor or not. We assert that, with the strategies as specified, $v^G_i(\mathcal{I})$ is indeed 0, for every $i$ and for every $\mathcal{I}$. That is, this form of purgatory is constructed so that its continuation value is 0: This is so because, once in $P$, $i$ is getting no favors. And the only way she can escape is by doing favors for players in state $G$ that are mandated by $S$. But if she does one of these favors for, say, $j$, $j$ randomizes between letting her back into $G$ and keeping her in $P$, with a probability that is set so that her expected benefit from getting back to $G$ just covers the cost of the favor she is called upon to do. In equilibrium, she does the favor; but she is indifferent between doing so or not. Since she is always

\[\text{The expected payoff (A\textsubscript{i}(S) - B\textsubscript{i}(S))/r} \text{ accrues as long as there is no “premature” descent by players into autarky. And in part 4 of the description of strategies, the prospect of autarky is raised. But this prospect is only realized if a player fails to follow the strategies outlined. More generally, from any starting point of the game (summarized by who is in G and who is in P), a player who is in G will receive the benefit of all favors for him that are mandated by S, as long as everyone sticks to the strategies. This player at worst must do all favors mandated by S; he may be able to do fewer, if some other players begin in state P.}\]
indifferent (between doing favors called for under $S$), her expected payoff is the same if she chooses not to do these favors; then she never does a favor, never gets a favor, and has value zero.

As for favors she might do while in P that are not mandated by $S$, part 6b says that $j$, the favor receiver, will definitely make the public declaration that keeps her in P; she receives no compensation for doing such a favor and so, of course, she doesn’t do it.

What about the public randomizations in parts 4, 6b, and 6c? In part 4 and 6c, we need $j$ to be indifferent between his two possible cheap-talk declarations, since the strategies call for him to use them both. (In part 4, circumstances dictate which declaration he uses; in part 6c, he employs a mixed [behavioral] strategy.) He weakly benefits by sending $i$ to purgatory in part 4 and by keeping $i$ in purgatory in part 6c, so to make him indifferent, if he sends $i$ to purgatory in part 4 or keeps $i$ in purgatory in part 6c, we mix with an outcome that is bad for him. In part 4, that outcome is autarky for all; we can’t use this device in part 6c for perfection considerations, but we can send $j$ to state P, which is just as good (that is, just as bad, for $j$). This then allows us to compute: In part 4, if $\mathcal{I}$ is the set of folks in P when $i$ fails to do a favor for $j$ that she was meant to do, and $j$ renounces her, the probability that we do not move to autarky is $v_j^G(\mathcal{I})/v_j^G(\mathcal{I} \cup \{i\})$. And in part 6c, where $i$ has done a favor mandated under $S$ for $j$ but $j$ has declared that $i$ is not restored to grace, the probability that $j$ is not sent to P is given by the same $v_j^G(\mathcal{I})/v_j^G(\mathcal{I} \cup \{i\})$, where now $\mathcal{I}$ is the set of people other than $i$ who are currently in P.

This leaves the public randomization in part 6b. The circumstances here are that $i$ is in P, $j$ is in G, and $i$ did a favor for $j$ that is not in $S_{ij}$. The declaration issued by $j$ is sure to keep $i$ in $P$, but since players other than $i$ and $j$ can’t distinguish between situation 6b and 6c (they are distinguished by whether $(x, y) \in S_{ij}$ or not, which only $i$ and $j$ know), what happens to $j$ under a public randomization if he says “$i$ is not restored to grace” must be the same in the two cases.

Once it is established (as we have done) that the continuation value of being in state P is zero, and once the various randomizing probabilities are set to get the indifferences indicated above, it is clear that the strategy array that has been described constitutes a perfect equilibrium that implements $S$. 

6. **Comparing ex-ante and ex-post cheap talk**

Propositions 1 and 3 tell us that, at least insofar as implementing time-homogeneous selections are concerned, everything that can be done with ex-ante cheap talk for a given $r$ can be done with ex-post cheap talk. More precisely, every selection $S$ that passes an obvious incentive-compatability test, Condition A, can be implemented as a perfect social equilibrium with ex-post cheap talk; since Condition B is required for ex-ante cheap talk to work, ex-post cheap talk seems superior.

We urge some caution in drawing this conclusion, on two grounds:

1. The equilibrium strategies in Proposition 1 are quite simple and straightforward; the off-path strategies

---

17 We could equally well, in part 4, move $j$ to P instead of moving everyone to autarky; this has the same impact on $j$’s incentives to issue his two possible cheap-talk declarations.
in Proposition 3 make sense if you stare at them long enough, but they cannot be called simple and straightforward.

2. As long as the incentive-compatibility constraint, Condition A, holds with strict inequalities, the equilibrium in Proposition 1 can be constructed with strict best responses for every player in every situation. It is part and parcel of the equilibrium in Proposition 3 that the favor receiver is indifferent whether or not to forgive a favor that is not done along the path of play, that (off the path of play) favor receivers use randomized strategies, that they are indifferent between allowing the favor giver back into a state of grace or not, and that a favor giver who is in state P is indifferent whether to do the favors called for under \( S \). (Suppose a favor giver in state G fails to do a mandated favor and then quietly warns the favor receiver, “If you denounce me and I go to P, I will not do any favors for you. This threat causes the equilibrium to collapse; it is essential that she does those favors, so that favor receivers is [weakly] better off with her in state P.) To the extent that all these indifferences make the equilibrium delicate, the equilibrium in Proposition 1 is better, when it works.

In this regard, it is possible, with ex-post cheap talk, to construct equilibria that implement some selections \( S \) without all the indifferences required in the equilibrium of Proposition 3. The first sort of indifference—a favor receiver who does not receive a favor along the path of play is indifferent in what he says—cannot be avoided; the argument given at the start of Section 5 shows this. But, by constructing other forms of punishment, one can create out-of-equilibrium strategies that have much more robust (that is to say, strict) best responses. These equilibria, however, require that there is a lot of “slack value,” in the sense that the left side of the inequality in Condition A is substantially less than the right-hand side; moreover, they require that inequalities of this sort hold where in place of the left-hand side we have the most expensive favors that a player might ever do, in the all-favors selection. In the on-line appendix, we provide details of some of these equilibria.

Of course, the ultimate test for which of ex-ante or ex-post cheap talk is employed is empirical. It is clear to us that both are employed in different contexts. The most we can say here is that, perhaps, this analysis suggests how the two might work, which can then be used to guide empirical investigation.

7. Concluding remarks: Extensions and variations

While Propositions 1 and 3 paint a nice theoretical picture, it remains to adapt these results and the stylized model we have studied to models that capture more of the institutional features of real-life models.

Time homogeneity and unilateral discretion

For one thing, we have focussed exclusively on the time-homogeneous implementation of selections \( S \). By this we mean, at least along the (equilibrium) path of play, the favors that are done are precisely those in \( S \), no more and no less. In both our equilibrium constructions, the only alternative to this, even
off the path of play, is autarky. But in other equilibrium constructions that are presented in the on-line appendix, we look at cases where, off the equilibrium path, other favors are done. An even broader study would consider equilibria in which the favors that are done along the path of play depend on the history of play up to that point in nontrivial ways.

In this regard, recall the brief mention of unilateral discretion back in Section 3. Even with no cheap talk by favor receivers, perfect social equilibria can be constructed in which the favor giver does not do all favors, but exercises her own discretion in which favors she chooses to do. In these equilibria, it is essential that the continuation payoff for the favor giver after she does a favor exceeds her continuation payoff if she doesn’t do a favor; were they the same, she would do no favors. This implies two things: First, what happens along the path of play cannot be time homogeneous. And, second, the favors she will choose to do will be determined entirely by their cost (that is, in any social equilibrium), since her continuation values can only depend on whether she does the favor or not. Hence if we look for such unilateral-discretion equilibria, considerations in the spirit of Condition B from Proposition 1 intrude. Unilateral discretion can only involve selections that satisfy Condition B. But, if one has a selection that satisfies Conditions A and B with strict inequalities, a folk-theorem without cheap talk can be shown: as the interest rate \( r \) approaches 0, equilibria can be constructed that asymptotically give the expected payoff of the selection. Conversely, if condition B fails for the selection \( S^U \), an anti-folk theorem along the lines of Proposition 2 can be proved. For details, see the on-line appendix.

Varying the informational endowments

The folk theorem just mentioned raises the general issue of information endowments: We have supposed throughout that \( i \) and \( j \) both know the precise values of cost and benefit of all \( i \)-for-\( j \) favors and that all the other players know when favor opportunities arise and whether the favor has been done. The folk theorem can be read as saying: If a selection \( S \) satisfies Conditions A and B with strict inequalities, and if the players are quite patient, then it is (asymptotically) unnecessary that the favor receiver of any favor know anything more than the general population of players, and the favor giver only needs to know the value of \( x \), her cost of doing the favor. Two other “informational variations” on our propositions constitute low-hanging fruit:

1. Suppose that \( i \) and \( j \) both know \( (x, y) \) for every \( i \)-for-\( j \) favor, but other players are generally unaware that a favor opportunity has occurred. The equilibria in Propositions 1 and 3 can still be made to work. In the equilibrium of Proposition 1, if an \( i \)-for-\( j \) favor opportunity arises for which \( (x, y) \in S_{ij} \), \( j \) can broadcast, “Player \( i \) has the opportunity to do me a favor, which I expect her to do.” It is necessary that other players can then see if the favor is done, but as long as they can see this, the equilibrium goes through. The key to this is that \( j \) is punished for a favor that \( i \) could do for her but does not do (and for which \( i \) is not excused); the same punishment would stop \( j \) from making the broadcast just suggested, if the favor is one that \( i \) will not do.
And for the equilibrium of Proposition 3: If an \(i\)-for-\(j\) favor opportunity arises, which \(i\) does not do, if \(i\) and \(j\) are in \(G\), and the favor should have been done (is in \(S\)), \(j\) can broadcast, “Player \(i\) had the opportunity to do a favor for me, should have done so, and didn’t do it.” Then we have a public randomization in which the two outcomes are that \(i\) alone goes to state \(P\), or both \(i\) and \(j\) go to \(P\); since this randomization makes \(j\) indifferent between talking and remaining silent, we can suppose that he speaks. Things are a bit trickier when \(i\) is in \(P\): If an \(i\)-for-\(j\) favor opportunity arises which \(i\) doesn’t do, \(j\) is perfectly happy (ex post) to let it ride; if \(i\) does a favor, the equilibrium requires that \(j\) make some declaration, although he would prefer not to, so we require that other players observe that \(i\) did the favor and it is \(j\)’s turn to speak.

2. And in the equilibrium of Proposition 1, it is not necessary that both \(i\) and \(j\) know the cost and benefit levels of each \(i\)-for-\(j\) favor; they only need to know whether or not the favor type \((x, y)\) is or is not in \(S\). For the equilibrium of Proposition 3, things aren’t quite so simple: Off the path of play, to achieve the correct randomization, a favor receiver must know the exact value of \(x\). But favor givers can get by with only knowing if a given favor is or is not part of the selection \(S\).

These are easy extensions; more interesting questions, which require serious analysis, concern how robust our constructions are to other variations in information endowments. A few simple results of this sort are given in the on-line appendix.

**Endogenous informational endowments**

In our analysis, the information endowments of the players are given exogenously. But it isn’t much of a leap from our models to the conclusion that parties involved in these types of transactions might want to invest in and then employ local information. This “investment in information” is, of course, something we see in the real world: Countries invest in embassies and diplomacy, trade missions, and multi-national trade forums, all of which serve these goals (and, of course, other goals.) As mentioned in the introduction, Toyota spends significant resources to understand suppliers’ cost structures and to have suppliers understand the manufacturing environment of Toyota, as well as on facilitating communication among its suppliers. Deans, department chairs, and (more generally) managers of all stripes are encouraged to “get to know” the personal concerns of the people they manage.

**Externalities**

We’ve looked here at the case where, when an \(i\)-for-\(j\) favor opportunity arises, only \(i\) and \(j\) are potentially affected directly by whether the favor is done or not. Our focus on this case is based on the premise that, when information is local, it is most likely to be held by the two parties most directly concerned with whether the favor is done. But in at least some of the applications we have in mind (e.g., organizational behavior in work settings), favors can generate externalities for the other parties in the “game.”
To say a bit more here about the application to work settings, imagine that we have an organization consisting of a collection of individuals. Instead of thinking of the random arrival of \(i\)-for-\(j\) favors, think of the random arrival of tasks for the various players \(i\). Each task, if done, has a cost \(x\) for the player \(i\) who does the task, and it generates a vector of positive benefits for all the other players. Imagine that some distinguished player, say player 1, who might be called “boss” or “department chair” or “dean” or “manager,” is able to tell, when a task-for-\(i\) arises, what is \(x\) and what is the vector of benefits for all the other players, including himself. (In addition, \(i\) must have this information.) Then, if we reinterpret a task-for-\(i\) as an \(i\)-for-\(1\) favor with externalities, we are back in the setting of our model (albeit with externalities). Note well, we do not require that the “referee” for which tasks should be done or should have been done—that is, player 1—is neutral to whether the task is done or not. Player 1 can receive external benefits if the task is done.

So it is interesting to ask, How do externalities affect our equilibrium constructions?

If we suppose that both \(i\) and \(j\) know which \(i\)-for-\(j\) favors ought to be done and which not (which, if favors generate externalities, could involve more than \(i\)’s costs and \(j\)’s benefits), then the equilibrium of Proposition 1 works nearly without change. The only modification is that in assessing whether Condition A holds for each player \(i\), one needs to compute the equilibrium net benefits that \(i\) gets in equilibrium, including net benefits from externalities from favors in which \(i\) is otherwise unconcerned.\(^{18}\)

Things are not so pleasant with the equilibrium of Proposition 3 (hence this is another reason not to dismiss ex-ante cheap talk): To give a sense of what can go wrong, suppose an \(i\)-for-\(j\) favor opportunity arises which \(i\) is meant to do according to \(S\), but which \(i\) doesn’t do. In the equilibrium, \(j\) should denounce \(i\), and \(j\) is willing to do so, because having \(i\) in state \(P\) is weakly better for \(j\). But suppose that some third party \(k\) does favors for \(i\) that generate particularly nice positive externalities for \(j\). If \(j\) sends \(i\) to \(P\), \(k\) stops doing favors for \(i\). So it may be worse for \(j\) to have \(i\) in \(P\) than in \(G\), which (of course) causes the equilibrium construction to collapse.

One of the alternative ex-post cheap talk equilibria we construct in the on-line appendix doesn’t fall afoul of these problems, at least in the case where all externalities are positive. See the on-line appendix for details.

**Transfer payments**

We began by ruling out transfer payments between the parties; since utility is transferable to some extent in the real world, we conclude with a few remarks on this point. If utility is transferable among the parties on a one-to-one basis, and absent the presence of externalities, the point of this paper is lost. When an \(i\)-for-\(j\) favor opportunity arises, where both \(i\) and \(j\) know the cost to \(i\) and the benefit to \(j\), one expects them to come to some mutually agreeable, bilateral deal. Assuming the cost to \(i\) is less than

---

\(^{18}\) And note in this regard the special case where player 1 is always treated as the favor receiver, because he is a specialist in assessing the costs and vector of benefits for each task opportunity. If player 1 announces, “This is a task that \(i\) ought to do,” and then \(i\) fails to do it, both \(i\) and player 1 should be punished. Readers can judge from personal experience if department chairs and deans are dealt with in this fashion. Of course, if player 1 has tasks to do other than keeping tabs on others, someone must monitor her.
the benefit to \( j \), \( j \) can offer to pay \( i \) her cost; or \( i \) can demand from \( j \) a payment equal to his benefit. If the cost to \( i \) exceeds the benefit to \( j \), there is no reason for the favor to be done.\(^{19}\) The issue becomes one of an immediate bargain being struck.

But, in some instances, monetary transfers are not possible and, even where they are possible, they may be inefficient; that is, it may cost an \( i \) more in utility than she can deliver to \( j \), in a direct transfer. Our assumption of no transfers is, of course, extreme. But it is not entirely unreasonable.

Adding together the possibility of transfer payments and externalities gives rich ground for further analysis, especially if the transfer payments can be made surreptitiously. Consider, in this regard, the equilibrium of Proposition 1. An \( i \)-for-\( j \) favor opportunity arises, that will cost \( i \) the amount \( x \) and provide benefits \( y < x \) to \( j \), but that generates further (positive) externalities for all the other players in amounts large enough so that, on efficiency grounds, this is a favor that ought to be done. If \( i \) can approach \( j \) before \( j \) broadcasts his ex ante message and offers him, say, \((x + y)/2\) if he will excuse this favor. \( j \) is better off accepting than refusing to issue the excuse. So, in particular, if we enlist the specialization in which some distinguished player, the department chair, say, is always in the position of favor receiver/referee, we begin to see a case (insofar as surreptitious transfers are possible) for constructing and enforcing sanctions against corrupt practices. (Or, we see why some decisions of this sort might be moved from the level of, say, a department chair—who may be more susceptible to under-the-table side payments of various sorts—to a decanal level.)

\[\text{Graduate School of Business, Stanford University}\]

\section*{Appendix}

\textit{The proof of Proposition 2.}

We want to show that the sum of the payoffs in any perfect social equilibrium (for a fixed value of \( r \)), when normalized, is uniformly bounded away from the normalized sum of the u-efficient payoffs. So fix some \( r \) and a perfect social equilibrium. The normalized sum of u-efficient payoffs is

\[
\sum_i \left[ \sum_{j \neq i} \lambda_{ij} A_{ij} - \lambda_{ij} B_{ij} \right] = r \sum_i \sum_{j \neq i} \sum_{n=1}^\infty E \left[ e^{-rT^n_{ij}} (y^n_{ij} - x^n_{ij}) 1_{\{x^n_{ij} < y^n_{ij}\}} \right],
\]

where \( T^n_{ij} \) is the random arrival time of the \( n \)th \( i \)-for-\( j \) favor and \((x^n_{ij}, y^n_{ij})\) is the cost-benefit vector for this favor. (If \( a_{ij} = 0 \), either omit \( j \) from the second sum for this \( i \) or take \( T^n_{ij} \equiv \infty \).) On the other hand, the normalized sum of payoffs in the equilibrium is

\[
r \sum_i \sum_{j \neq i} \sum_{n=1}^\infty E \left[ e^{-rT^n_{ij}} (y^n_{ij} - x^n_{ij}) 1_{\{i \text{ does this favor for } j\}} \right].
\]

\(^{19}\) Think of the circle and a case where all \( i \)-for-\( i+1 \), for \( i = 1, 2, \ldots, I - 1 \), have cost 1 and benefit 2, while the \( I \)-for-1 favors have cost 2.1 and benefit 2. You might at first think that we need for this last, \( u \)-inefficient favor to be done, to keep 1 happy and willing to do favors for 2, and so on. But with transferable utility, this simply isn’t so. Player 2 pays player 1 for 1-for-2 favors, which is all the compensation 1 needs.
where the expectation being taken involves the exogenously determined random arrival times of favors, the exogenously determined randomly benefit–cost vectors for each favor, and also the endogenously determined strategies being employed in the perfect social equilibrium under investigation. Hence the difference between the normalized sums of the u-efficient payoffs and the equilibrium payoffs is

\[ r \sum_i \sum_{j \neq i} \sum_{n=1}^{\infty} E \left[ e^{-rT_{ij}^n}(y_{ij}^n - x_{ij}^n)\left[1\{x_{ij}^n < y_{ij}^n\} - 1\{i \text{ does this favor for } j\}\right]\right]. \quad (A1) \]

(We emphasize that the occurrence and timing of favor opportunities is exogenous—that is, unaffected by the strategies the players employ—so writing the difference this way is legitimate.) Note that each term inside the larger square brackets is nonnegative: If \( y_{ij}^n > x_{ij}^n \), the first indicator function is 1, so the term inside the smaller square brackets is either 1 or 0; if \( y_{ij}^n < x_{ij}^n \), the first indicator function is 0, so the term inside the smaller square brackets is either 0 or \(-1\). Therefore, overall difference is always nonnegative (of course), and we underestimate the difference if we look only at some of the terms in the triple summation.

The proposition posits that, for some ordered pair \( i \) and \( j \), there are four cost-benefit vectors \((x(k), y(k))\) (in the support of the distribution of cost-benefit vectors for \( i \)-for-\( j \) favors) such that \( y(1) > x(1) > x(2) > y(2) > y(3) > x(3) > x(4) > y(4) \). Refer to an \( i \)-for-\( j \) favor with cost-benefit vector \((x(k), y(k))\) as a favor of type \( k \), for this ordered pair, and we let \( p_k \) be the probability that, in an \( i \)-for-\( j \) favor, the cost-benefit vector makes this a favor of type \( k \).

Now go back to the difference in \((A1)\), and examine the term for this specific \( i \) and \( j \) (ordered) pair. That is, we are looking at

\[ r \sum_{n=1}^{\infty} E \left[ e^{-rT_{ij}^n}(y_{ij}^n - x_{ij}^n)\left[1\{x_{ij}^n < y_{ij}^n\} - 1\{i \text{ does this favor for } j\}\right]\right]. \quad (A2) \]

The occurrence and timing of \( i \)-for-\( j \) favors is independent of the cost-benefit vectors for those favors, so each time there is an \( i \)-for-\( j \) favor opportunity, there is probability \( p_k \) that it is of type \( k \). Note that for favors of types 1 and 3, the cost is less than the benefit, so they should be done to achieve u-efficiency, while favors of types 2 and 4 should not be done.

We now demonstrate the following: In any perfect social equilibrium, at every occurrence of an \( i \)-for-\( j \) favor, one of the following four conditions must hold: \( j \) does not do the favor if it is of type 1, or \( j \) does not do the favor if it is of type 3, or \( j \) does the favor if it is type 2, or \( j \) does the favor if it is of type 4. Hence, for

\[ K^1 = \min \{p_1[y(1) - x(1)], p_2[x(2) - y(2)], p_2[y(3) - x(3)], p_4[x(4) - y(4)]\}, \]

Proposition 2 is established, for \( K = a_{ij}K^1 \).

The proposition posits that \( j \) can only issue one of two cheap-talk messages, which we denote by \( A \) and \( B \). We use \( v_{AF} \) for \( i \)'s continuation value if \( j \) issues message \( A \) and \( i \) does the favor, \( v_{AN} \) for
i’s continuation value if j says A and i does not do the favor, and similarly for $v_{BF}$ and $v_{BN}$. And we use $u_{AF}$, and so forth, for j’s continuation values.

Suppose at some point in the game, with some public history, at some time, an i-for-j favor opportunity arises.

1. Perhaps i’s strategy at this point is to refuse to do the favor regardless of what j says. If so, then the italicized assertion is true. So suppose this is not true; i will do the favor with positive probability if (at least) the message is $A$. (The choice of $A$ is, of course, without loss of generality.)

2. But if i does the favor with positive probability when the favor is of type 1, we know that $v_{AF} - x(1) \geq v_{AN}$. Since $x(1) > x(2) > x(3) > x(4)$, this implies that $v_{AF} - x(i) > v_{AN}$, which means that, if the favor is of type 2, 3, or 4, and j says A, i is certain to do the favor.

3. And by a similar argument, if i will do the favor with positive probability when j says B and the favor is of type 2 or type 3, then i will do the favor with probability 1 when j says B and the favor is of type 4. But if i will do the favor with probability 1 regardless of the message when the favor is of type 4, then the italicized assertion is true.

4. So the only way the italicized assertion could fail to be true is if i will do the favor with certainty for favors of type 2 and 3, when j says A, and i refuses to do the favor with certainty for favors of type 2 and 3, when j says B. Now look at j’s incentives. When the favor is of type 2, if j says A, he gets $y(2) + u_{AF}$, while if he says B, he gets $u_{BN}$. If $u_{BN} < y(2) + u_{AF}$, then j will always say A when the favor is type 2, the favor is done, and the italicized assertion is true. But if $u_{BN} \geq y(2) + u_{AF}$, then $u_{BN} > y(3) + u_{AF}$, j will say B when the favor is type 3, a favor of type 3 is not done, and the italicized assertion is true.

The italicized assertion is true, completing the proof of Proposition 2.

It should be clear that the argument just given relies heavily on the assumption that the favor receiver is allowed to send only one of two messages. We conjecture that, if the favor receiver is allowed as many messages as there are $(x, y)$ pairs in the support of their distribution, asymptotic u-efficiency can be achieved with ex ante communication, as long as $A_i(S^U) > B_i(S^U)$ for all i. Indeed, we conjecture that if a selection $S$ satisfies $A_i(S) > B_i(S)$ for all i, then with a sufficiently rich language of ex ante cheap talk by favor receivers, perfect social equilibria for small interest rates $r$ can be constructed, which, as $r \to 0$, have normalized payoffs that approach the normalized payoffs of $S$. Suppose momentarily that players can engage in utility transfers. Having seen the cost-benefit vector $(x, y)$, the favor receiver says to the favor giver, “If you do the favor for me, I will pay you x,” if $x < y$. If $x > y$, the favor giver makes no offer. The favor giver, then, is indifferent between doing the favor or not when $x < y$, so we can assume she will do it. The favor receiver, of course, will happily pay $x$ for a favor of value $y$, if $x < y$, but will be unable to offer any payment sufficient to induce i to do the favor, if $x > y$.

We do not have transferable utility. But, following the general techniques of Fudenberg, Levine, and Maskin (1994), when $r$ is close to zero, the “continuation” of the game can be used in lieu of transferable utility, as long as all players are aware of the “deal” struck between i and j and the appropriate full-
dimensionality assumption is met. There are details to be checked, of course, so we only call this a conjecture. But it seems to us to be a fairly safe conjecture.

References


