DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

RATIFIABLE MECHANISMS: LEARNING FROM DISAGREEMENT

Peter C. Cramton
Yale School of Management

Thomas R. Palfrey
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 731

April 1990

Ratifiable Mechanisms: Learning from Disagreement

by

Peter C. Cramton and Thomas R. Palfrey*

April 1990

Abstract

In a mechanism design problem, participation constraints require that all types prefer the proposed mechanism to some status quo alternative. If the payoffs in the status quo depend on strategic actions based on the players' beliefs, then the inferences players make in the event someone objects to the proposed mechanism may alter the participation constraints. We include this possibility for learning from disagreement by modeling the mechanism design problem as a ratification game in which privately informed players simultaneously vote for or against the proposed mechanism. We develop and illustrate a new concept, ratifiability, that takes account of this inferencing problem in a consistent way. Requiring a mechanism to be ratifiable can either strengthen or weaken the standard participation constraints that arise in mechanism design problems.

*Yale School of Management and California Institute of Technology, respectively. We thank Jeff Banks, Steve Matthews, Preston McAfee, Joel Sobel, and numerous seminar participants for valuable comments. We are grateful to the National Science Foundation for support.

Mechanism design is a powerful theory for studying incentive problems in settings where privately informed decision-makers have conflicting interests. In a typical application of the mechanism design approach, the analyst is able to characterize the set of outcomes that are attainable by the agents, recognizing each agent's voluntary participation and incentive to misrepresent information that is privately known. In addition, it is often possible to characterize "optimal" mechanisms – mechanisms that are efficient in either an ex ante or interim sense.

Despite these significant virtues, one can argue that the mechanism design approach identifies too large a set of attainable outcomes, because it allows the agents to make unreasonable commitments, both before and after the selection of a mechanism. Commitment becomes an issue if information is leaked during the selection or implementation of a mechanism. The information leakage alters the incentive problem faced by the agents, creating opportunities for renegotiation. An inability to commit to not using leaked information typically reduces the set of attainable outcomes, because additional incentive constraints must be satisfied. Several authors have focused on the commitment problem face by agents as a result of information leakage during the implementation of a mechanism. See, for example, Ausubel and Deneckere (1987), Caillaud and Hermalin (1989), Cramton (1985), and Green and Laffont (1984, 1985).

Others have addressed the related problem of information leaked during the process of selecting a mechanism. For example, Myerson (1983) and Maskin and Tirole (1988, 1990) analyze the problem a privately informed principal has in selecting a mechanism recognizing that the selection may reveal information to the subordinates. Ideally, the principal would like to condition her choice of mechanism on her private information, but to do so would reveal information and hence make the chosen mechanism invalid, assuming it is based on the prior beliefs of the subordinates.

One way around this problem of information leakage during the selection of a mechanism is to assume that the agents select the mechanism in the ex ante stage, before they have their private information. But even if selection is done by uninformed agents, once the agents learn their private information – the interim stage – they may have an incentive to renegotiate to a different mechanism, as Holmström and Myerson (1983) demonstrate. Again, if the agents are unable to commit to not renegotiating, then the set of attainable outcomes is further constrained.

Here we focus on each agent's decision to participate in the mechanism and explore the possibility that a refusal to participate may reveal information. To isolate this information leakage problem from the others, we assume:

- The proposed mechanism is selected by an uninformed third-party, so no information is revealed by its selection.
- The agents can commit to not renegotiating to an alternative proposal at the interim stage, thus avoiding the durability problem discussed in Holmström and Myerson (1983) and Crawford (1985).
- The agents voluntarily decide to participate in the mechanism at the interim stage. If participation is unanimous, then the agents are bound by the mechanism; otherwise, they play a *status quo game*, possibly with altered beliefs. In either case, the agents are ultimately committed to the mechanism or the status quo and cannot renegotiate during implementation.

A special case of this model is widely found in the mechanism design literature, where it is typically assumed that nothing is learned from selection, that the agents can commit to the selected mechanism, and that agents voluntarily decide to participate after learning their private information, so interim individual rationality is the relevant set of participation constraints. It is important that an uninformed third-party selects the mechanism, rather than the agents in the ex ante stage, since then presumably the agents would be able commit to the mechanism in the ex ante stage, so only ex ante individual rationality, rather than interim individual rationality, would be required. Our commitment assumption, while strong, is not unreasonable in an environment where contracts can be enforced.[1]

We allow the possibility for learning from the agents' participation decisions by modeling the ratification process as a two stage game: a voting stage in which each player simultaneously votes for or against the proposed mechanism followed by a implementation stage in which either the proposed mechanism or the status quo is adopted, depending on whether or not the proposed mechanism was unanimously ratified in the voting stage.[2] A mechanism is *ratifiable* relative to the status quo if unanimous ratification is a sequential equilibrium of the two stage game, where beliefs following disagreement are required to satisfy consistency conditions similar to those proposed by Farrell (1985) and Grossman and Perry (1986).

In many mechanism design settings, learning from disagreement is not an issue, since the status quo outcome is not affected by a change in the players' beliefs. For example, in bilateral trade, a natural status quo would be no trade (Myerson and Satterthwaite, 1983). But in many other situations (and perhaps even in the bilateral trade case), equilibrium play in the status quo game may depend on the players' revised beliefs, *conditioned on the failure to ratify the proposed mechanism*. If this is the case, then ratification gives a player an opportunity to signal information, which in turn may affect the desirability of playing the status quo game. Hence, an agent's

[1] Renegotiation becomes a problem if the parties end up in a situation where everyone agrees to change the mechanism. This state, however, can apparently be avoided if communication is limited or if the agents include in their initial contract a clause requiring all parties to pay large sums of money to a third-party in the event the contract is renegotiated. A court interested in efficient contracting should enforce such clauses. The argument is not so simple, however, since the third-party may agree to tear up the contract for a small fee, since if the contract is not torn up the parties will never renegotiate and so the third-party will get nothing. The argument, therefore, rests on the assumption that the third-party is not susceptible to such bribes.

[2] Maskin and Tirole's (1988,1990) analysis of the informed principal problem also has a ratification stage, but our models are quite different. In Maskin and Tirole an uninformed agent chooses to ratify the mechanism. Information is revealed from the informed principal's selection of a mechanism, rather than the ratification process. In our model, an uninformed principal selects the mechanism and then informed agents choose to ratify. Information is revealed from the ratification process, but not from the choice of mechanism.

participation decision depends on what he expects others to infer from his veto of the proposed mechanism.

We now give three examples of mechanism design problems where learning during the ratification process is relevant.

Cartels (Cramton and Palfrey, 1990). The firms in an industry wish to form a cartel. Each firm has private information about its marginal cost, as in Roberts (1985). Under the proposed (monopoly) mechanism, the firms announce their costs, the firm with the lowest cost produces the monopoly level of output, the others produce nothing, and the monopoly revenue is split among the firms based on the vector of reported costs so as to elicit truthful reports. If the proposed mechanism is not unanimously ratified, a status quo game consisting of (Bayesian) Cournot competition ensues: each firm simultaneously selects output to maximize profit given its own costs and its belief about the costs of the other firms and how much they will produce. But what beliefs should the firms hold in the event of disagreement? These beliefs are important, because they determine a firm's payoff in the status quo, and hence whether it should ratify the proposed mechanism. Individual rationality critically depends on what is inferred from a (possibly unexpected) veto of the proposed monopoly mechanism.

One possibility, consistent with the standard mechanism design approach, is that nothing is learned from disagreement. With two firms, if the proposed mechanism has been designed to be incentive compatible and interim individually rational (assuming nothing is learned), then all types of firms have an interest in ratifying the mechanism, so disagreement is a zero-probability event. Passive beliefs (no learning) will support the proposed mechanism as a sequential equilibrium in this ratification game. But other beliefs are possible as well.

We argue that passive updating is implausible. The ratifying firm, surprised by the disagreement of the other, should try to rationalize the deviation by identifying a set of types (called the *veto set*) that could stand to benefit from vetoing. The veto set is credible if those in the veto set prefer the status quo game to the mechanism, and those not in the veto set prefer the mechanism to the status quo game, where the status quo payoffs are calculated at an equilibrium with the updated belief that the vetoer is in the veto set. In this way, we restrict beliefs about a vetoing firm to be a credible veto set of types (if one exists).

In our cartel example, it is low-cost firms that benefit the least from participating in the cartel. Might a low-cost firm stand to gain by vetoing if by doing so it signalled that it had a low cost? If the others believe that the vetoing firm has a low cost, then they expect this firm to produce a lot and so they optimally respond by reducing their output, thus increasing the status quo payoff to the vetoing firm. Since low-cost firms gain the least from participation, this increase in the status quo payoff may be sufficient for them to prefer the status quo to the mechanism, but high-cost firms (who gain a great deal from the mechanism) would still prefer the mechanism to the status quo even if by vetoing a high-cost firm is able to convince the others that it is a low-cost firm. Here learning from disagreement has the effect of strengthening the participation constraints by improving the status quo payoffs for the vetoer, and therefore reduces the set of attainable outcomes.

Litigation (Spulber, 1988 and Spier, 1988). A plaintiff and defendant are engaged in pretrial negotiation. The plaintiff has private information about her level of damages and the defendant has private information about his liability. They would like to settle their dispute without a trial, because the trial is costly to both. The proposed mechanism specifies a settlement amount and a probability of going to court as a function of the reports of private information. The status quo is going to court *if* the plaintiff believes bringing the case to court is profitable. If the case does go to court, each player's private information is revealed during the costly discovery process and then the court awards to the plaintiff her damages times the degree of liability of the defendant.

Notice that the defendant's status quo payoff depends on the plaintiff's decision to take the case to court, which in turn depends on the plaintiff's belief about the liability of the defendant. If the plaintiff's damages are sufficiently small or she believes the defendant is not sufficiently liable, then the plaintiff prefers not to take the case to court, since the cost of discovery and trial is greater than the expected award.

What type of defendant gains the least from participating in the mechanism? If the proposed mechanism is incentive compatible and does not suggest going to court for plaintiffs with damages that are too small to make going to court profitable, then the least-liable type of defendant gains the least from participating in the mechanism relative to the status quo. Vetoing the mechanism may be a credible signal of low liability, since signaling low liability reduces the probability that the plaintiff will take the case to court and hence raises the defendant's status quo payoff. So long as defendants with high liability still gain more from the mechanism than the improved status quo, the low liable types form a credible veto set. Learning from disagreement in this example would have the same effect as in the cartel example: the learning improves the status quo payoffs for the vetoer, and so the set of attainable outcomes is reduced.

Arbitration. A buyer and a seller are negotiating a price for an object owned by the seller. The buyer has private information about his reservation price, as does the seller. A mechanism specifies a trading price and a probability of trade as a function of the private information of the two parties. The status quo is either no trade or arbitration *if* both agree to arbitrate. Arbitration is a costly verification procedure similar to the courts in the litigation example. The arbitrator, through costly investigation, verifies the traders' reservation prices, and then has them split any gains from trade equally. Incentive compatibility typically implies that the highest valuation seller and the lowest valuation buyer gain the least from participating in such a mechanism. Can vetoing be a credible signal of a high value for the seller? In this case, signaling a high type is bad for the seller in the status quo game, since it reduces the probability that the buyer will be willing to bear the cost of arbitration. A low-value seller would have no interest signaling a high type by vetoing, since the low type benefits more from the mechanism than a high type. Hence, unlike the last two examples, here learning from disagreement has the effect of reducing the status quo payoffs for the vetoer, and thereby expanding the set of individually rational mechanisms.

The central feature of these examples is that the outcome of the status quo game depends on actions that the agents take based on their beliefs about private information. Thus, what is learned in the event that someone vetoes the mechanism affects the vetoer's status quo payoff and therefore the agent's decision to participate. Individual rationality, then, depends on what is learned from disagreement. In order to answer the question, "What are the relevant individual rationality constraints?", we must first answer the question, "What beliefs should the agents have following disagreement?"

Section 1 presents the general model and definitions. Section 2 applies the general model to the cartel problem studied in Cramton and Palfrey (1990). The cartel example serves to illustrate that even ex post efficient mechanisms that are feasible, incentive compatible, and individually rational without learning may be eliminated by refinements that place reasonable restrictions on beliefs following disagreement. Section 3 compares the implications of our model of learning from disagreement with several alternatives from the literature on equilibrium refinements.

1 Ratifiable Mechanisms

Consider a mechanism design problem with n individuals, indexed by $i \in N = \{1, \ldots, n\}$, that must make a decision $d \in D$. Each individual i has private information $t_i \in T_i$, representing a realization of all of i's information that is not common knowledge. For simplicity, we assume that types are independent: t_i is drawn from the distribution F_i, independently of t_j for $j \neq i$. Player i's

ex post utility $u_i : D \times T \rightarrow \mathbb{R}$ depends on the decision and vector of types $t = (t_1, \ldots, t_n) \in T = T_1 \times \cdots \times T_n$. A decision rule $\delta : T \rightarrow D$ maps each vector of types into a decision. The decision rule δ is implemented as a direct mechanism: the players simultaneously report their types t', possibly dishonestly, and then the decision $\delta(t')$ is adopted. An uninformed third-party, the mechanism designer, must propose a particular decision rule, recognizing incentive and participation constraints. Incentive constraints are dealt with by requiring that the decision rule δ be incentive compatible. $\delta(t)$ is incentive compatible if, in the direct mechanism, honesty is a best response for each player given that the others are honest. By the revelation principle, there is no loss of generality in requiring that the designer propose an incentive compatible decision rule. In what follows, δ will always refer to an incentive compatible decision rule. Participation constraints are handled by requiring that every type of every player prefers to participate in the mechanism δ than play the *status quo game* G.

In the standard mechanism design approach, the participation constraints are quite simple, because the payoffs in the status quo game, the alternative to participation, do not depend on strategic actions. Often the status quo simply specifies a constant value, say U^o. In that case, interim participation constraints are that the interim utility from the mechanism for each type of each player be at least U^o.

Here we allow for a status quo that depends on strategic actions. Participation constraints are considered by analyzing a two-stage ratification game. In the first stage, the players simultaneously vote to ratify the proposed mechanism δ. In the second stage, δ is implemented if it is unanimously ratified, otherwise the status quo game G is played. Player i, in deciding whether to veto the mechanism δ, must be aware that the others' beliefs about i might change as a result of the veto, and this change in belief might alter i's payoff in the status quo game.

Furthermore, the expected payoff to i in G if i vetoes depends upon the equilibrium that will be played in G under the belief generated by the veto. What we attempt to do here is to link together i's decision to ratify or veto with rational expectations about the outcome in the continuation game G.

In particular, suppose that if i vetoes δ then the others believe that i's type t_i is in some *veto set* $V_i \subseteq T_i$, $V_i \neq \emptyset$.[3] The status quo game is then played with the original beliefs, except all players other than i update their beliefs about player i's type to be $F_i(\cdot|V_i)$, where $F_i(\cdot|V_i)$ denotes the distribution F_i conditioned on the event $V_i \subseteq T_i$. Let $\sigma(V_i)$ denote an equilibrium of G with these new priors, and define $U_i^G(t_i, V_i, \sigma(V_i))$ to be the interim payoff to i in the equilibrium $\sigma(V_i)$ in G when i vetoes and the others infer i's type is in V_i. Finally, let $U_i^\delta(t_i)$ be t_i's interim payoff in the mechanism δ, and define $U_i(t_i, V_i, \sigma(V_i)) = U_i^\delta(t_i) - U_i^G(t_i, V_i, \sigma(V_i))$ to be t_i's net benefit from δ, relative to G, if a veto by i results in the inference that $t_i \in V_i$ and $\sigma(V_i)$ is played in the continuation game. Let $V = \{V_1, \ldots, V_n\}$ be a collection of veto sets, one for each player, and let $\Sigma(V) = \{\sigma(V_1), \ldots, \sigma(V_n)\}$.

Definition. δ is *individually rational relative to the status quo G* if there exists V and $\Sigma(V)$ such that for each i and each t_i, $U_i(t_i, V_i, \sigma(V_i)) \geq 0$.

Our definition is the standard definition of individual rationality, but now the payoff in the status quo can depend on what is learned from i's veto of the mechanism. Our interest in this definition stems from the following fact.

[3]The independence of the players' types is used here, since the veto set V_i when i vetoes is the same for all players. This simplification would not be possible with dependent types. Although convenient, the assumption of independence of types is far from innocuous in mechanism design as is demonstrated by Crémer and McLean (1988) in an auction environment.

Proposition 1. *If δ is incentive compatible and individually rational relative to G, then unanimous ratification of δ followed by truthful revelation in the direct mechanism δ is a sequential equilibrium in the ratification game.*

Proof. If in equilibrium all types of all players ratify δ, then Bayes' rule implies that the prior beliefs are unchanged when the direct game constructed to implement δ is played. Since δ is incentive compatible, truthful revelation in this game is a best response following unanimous ratification. Consider any player i and $j \neq i$. Let j's belief if i deviates by vetoing δ be $F_i(t_i|V_i)$, and let all players follow $\sigma(V_i)$ in the continuation game. Given this, a veto by i in the ratification stage is unprofitable for any type t_i, since the fact that δ is individually rational relative to G implies that $U_i(t_i, V_i, \sigma(V_i)) \geq 0$. ■

The proposition above is not an "if and only if" statement, because our definition of individual rationality relative to the status quo game G places an additional restriction on beliefs off-the-equilibrium-path beyond what is required for a consistent assessment in a sequential equilibrium. Namely, j's belief about i following i's veto is the conditional distribution $F_i(t_i|V_i)$. Our assumption, however, that any two players j and k have the same belief about i after i's veto is required in a sequential equilibrium, because types are independent. This follows, since for the beliefs to be consistent they must be the limit of beliefs formed by a sequence of totally mixed strategies.

1.1 Credible Veto Sets and Ratifiability

We now explore further restrictions on beliefs in the ratification game, based on a refinement proposed by Grossman and Perry (1986). In particular, we suppose that if i vetoes the mechanism, then the others will try to rationalize the veto by inferring that t_i is in a particular veto set V_i, such that all types in V_i benefit from the veto, but those not in V_i prefer the mechanism. If it is possible to rationalize the veto in this way, then the mechanism is not individually rational in the sense that types in V_i can do better by vetoing.

Definition. A nonempty subset V_i of T_i is a *credible veto set* for i relative to the decision rule δ and status quo game G if there exists some $\sigma(V_i)$ such that:

(i) $U_i(t_i, V_i, \sigma(V_i)) \leq 0$ for all $t_i \in V_i$, and

(ii) $U_i(t_i, V_i, \sigma(V_i)) \geq 0$ for all $t_i \notin V_i$.

If V_i is a credible veto set, then the others can rationalize i's veto by believing i's type is in V_i. If for every credible veto set there is some type t_i that strictly gains by vetoing, then δ is not ratifiable relative to G.

Definition. An incentive compatible decision rule δ is *ratifiable against* G if for all i either

(i) there does not exist a credible veto set, or

(ii) there exists a credible veto set V_i such that $U_i(t_i, V_i, \sigma(V_i)) = 0$ for all $t_i \in V_i$.

If after a veto the players' beliefs are restricted to credible veto sets when one exists, then we should require that the mechanism be ratifiable. For any mechanism that is not ratifiable, regardless of the credible veto set, there is some type of player that strictly benefits from vetoing. This definition is in the spirit of perfect sequential equilibrium (Grossman and Perry, 1986).

Remark 1. It is certainly possible that δ and G are such that no credible veto sets exist. In this case, δ is ratifiable, the participation constraints are not binding, and the beliefs following disagreement are indeterminate (and inconsequential). However, in many settings, the participation constraints are binding, provided the incentive problem is severe enough and the mechanism designer chooses δ to be optimal in some sense. Such a mechanism typically is associated with the existence of a

credible veto set for each i with the property that all types in V_i are indifferent between ratifying and vetoing. In this case, the beliefs following disagreement are required to put all the weight on those types that gain the least from participating in the mechanism δ relative to G.

Remark 2. A well-known difficulty with the perfect sequential equilibrium (PSE) concept is that PSE sometimes fail to exist. Because of the similarity of our notion of ratifiability to PSE, we must address the possibility that for some G, the set of ratifiable mechanisms relative to G may be empty. This, however, is never the case. Let $\sigma(T_i)$ be any equilibrium to G under the prior beliefs. Let δ be the decision rule generated by $\sigma(T_i)$. Then δ is always ratifiable against G, since for each i, T_i is a credible veto set with $U_i(t_i, T_i, \sigma(T_i)) = 0$ for all $t_i \in T_i$.

Remark 3. It is possible that there is more than one way to rationalize a veto by player i. In that case, there will be more than one credible set for i. In fact, it is possible that for one of i's credible sets, say V_i, $U_i(t_i, V_i, \sigma(V_i)) = 0$ for all $t_i \in V_i$ but for another of i's credible sets, say V_i', $U_i(t_i, V_i', \sigma(V_i')) < 0$ for some $t_i \in V_i'$. In this case, the mechanism would be ratifiable, as long as appropriate credible veto sets could be found for all $j \neq i$, even though there existed a credible veto set for i, V_i', that would make some i-types better off under the status quo than under the proposed mechanism. This problem of multiple credible veto sets will be addressed later in the paper. The issue is whether part (ii) of the definition of ratifiability should be required to hold for all credible veto sets. This, however, is not an issue if there is at most one credible set for any i.

1.2 Strong Ratifiability

As pointed out in Grossman and Perry (1986), there is a subtle difference between their definition of PSE and a related equilibrium refinement proposed by Farrell (1985). Farrell's refinement, *neologism-proof*, differs from PSE because it requires an equilibrium to be supported by all, rather than one, credible updating rules for rationalizing observations off the equilibrium path. If neologism-proofness is applied in our framework of ratifiability instead of PSE, we get a strengthening of our definition of ratifiability.

Definition. An incentive compatible mechanism δ is *strongly ratifiable* against G if for all i either

(i) there does not exist a credible veto set, or

(ii) for every credible veto set V_i, $U_i(t_i, V_i, \sigma(V_i)) = 0$ for all $t_i \in V_i$.

In other words, if δ is strongly ratifiable against G, then there does not exist a credible veto set for any player such that some type in that veto set strictly prefers the status quo to the mechanism. Any mechanism that is strongly ratifiable is ratifiable because part (ii) of the definition must hold for all credible sets, rather than just one.

In contrast to the definition of ratifiable, it may be that the status quo G is not strongly ratifiable against itself (or, more precisely, G is not strongly ratifiable against a decision rule produced at some equilibrium of G). There may be a credible set of types $V_i \neq T_i$ such that at least one type t_i strictly prefers an equilibrium in the status quo game with the revised beliefs V_i relative to the status quo with beliefs T_i. Such a game would not be impervious to allowing players to make binary preplay announcements ("Veto" or "Ratify") that may communicate information about their types. Thus, we can interpret the idea of a mechanism being strongly ratifiable against itself as permitting a special sort of preplay communication before G is carried out.[4] In this way we see that strong ratifiability implies some degree of cheap-talk proofness.

[4] Matthews, Okuno-Fujiwara, and Postlewaite (1989) provide an interesting analysis of pre-play communication in Bayesian games that is relevant here. As in Farrell (1985), their pre-play communication is cheap talk (it does not affect payoffs directly) but they allow a richer structure. In particular, different deviant types may send different messages. The communication in our model is quite different, since it stems from the binary decision to ratify. Communication in our model is not cheap talk and naturally splits the set of types into two subsets – those that vetoed and those that ratified the mechanism.

2 Ratifiability of Cartel Agreements with a Cournot Threat

We now turn to an example from Cramton and Palfrey (1990) that illustrates how learning from disagreement can affect the set of ratifiable mechanisms. Whether learning from disagreement strengthens or weakens the participation constraints depends in general on the particular mechanism δ. In our example, even though the status quo G is fixed, participation constraints are sometimes strengthened by learning from disagreement and sometimes weakened depending on the proposed mechanism δ.

Two firms produce in an industry. Each firm's marginal cost c_i is private information and drawn independently from the uniform distribution on [0,1]. Inverse demand is $p(q_1, q_2) = 1 - q_1 - q_2$, where q_i is the production of firm i. Both firms seek to maximize their interim profit. The status quo G is Bayesian Cournot competition: each firm i chooses q_i to maximize its expected profit given its cost c_i and its belief about j's production. A mechanism δ specifies how much each firm produces and how the industry revenue is divided between the firms as a function of their reported costs. We consider two different mechanisms. The first, joint monopoly, illustrates that learning from disagreement can strengthen the participation constraints – although joint monopoly is attainable without learning, it is not strongly ratifiable. The second, a minimum quantity restriction, demonstrates that learning from disagreement can have the opposite affect – although no minimum quantity restriction is individually rational without learning, it is ratifiable.

2.1 Joint-Monopoly Mechanisms

Let δ be the joint monopoly outcome: the lowest-cost firm produces the monopoly output and the other produces nothing with the revenue divided in such a way that δ is incentive compatible. Using the standard mechanism design approach, one can show that it is possible to split up the revenue in such a way as to satisfy interim individual rationality assuming nothing is learned from disagreement (that is, the status quo of Cournot competition is played with the prior beliefs). In this mechanism without learning, the worst-off type, $\hat{c}$ = 0.2200, expects to get 0.0999 in profits from the joint-monopoly mechanism and only 0.093 in the Cournot game.

But is the joint-monopoly mechanism ratifiable against Cournot competition? To answer this question, we need to ask: what should firm j think (and consequently how much should it produce) if firm i vetoes the monopoly mechanism? Does there exist a set of types V_i (a credible veto set) that can make the following speech:

> "I voted against this mechanism because my type is in V_i. If you believe me and I am telling you the truth, then my payoff in the Cournot game is better than what I get in the mechanism. Moreover, if my type is not in V_i, I would get a strictly higher payoff from the mechanism than from the Cournot game in which you believe my type is in V_i, and so I would not want to vote against the mechanism. Hence, you should believe me."

If such a credible veto set exists (and there is no other veto set satisfying (ii) of the definition of ratifiability), then the monopoly mechanism is not ratifiable. This turns out to be the case, as we demonstrate below.

The intuition for why a credible veto set exists here is straightforward. In the monopoly mechanism, high-cost firms gain a great deal by participating, whereas low-cost firms gain little. By vetoing the monopoly mechanism, a firm sends a credible signal that it has a relatively low cost. Since a firm with a low cost produces a relatively large amount in the Cournot status quo, the other firm will respond optimally by producing less in G than it would with its prior beliefs. This reduction of output by the other firm increases the profit to the vetoing firm in the status quo – enough, in fact, to make a vetoer with sufficiently low cost prefer the Cournot outcome to the monopoly mechanism. A low-cost firm's signal is credible, since a high-cost firm still does better in the monopoly mechanism, despite the improved status quo that results from vetoing.

Proposition 2. *The monopoly mechanism is not ratifiable against the status quo of Cournot competition.*

Proof Sketch. The idea behind the proof is that a mechanism is ratifiable only if the ratifier's expected output is sufficiently high in the status quo game following a veto. As the ratifier's expected output increases in the Cournot game, the vetoer's output in the Cournot game declines. In fact, it can be shown that for any beliefs following a veto, the subsequent equilibrium in the Cournot game will have the vetoer produce if and only if its cost is less than some amount. For future reference, denote such a critical cost level for the vetoer by $c_v(V)$, where V denotes beliefs the ratifier has about the veto set. We then show that whenever V is such that $c_v(V) \leq .8515$ then the vetoer will be worse off in the subsequent Cournot game that it would be in the monopoly mechanism. Finally, we show that the unique credible veto set, V, is the interval [0,.444] which results in a value of $c_v(V) = .89 > .8515$. See the appendix for details. ■

Learning in this case has the effect of strengthening the participation constraints. If a veto resulted in passive inferences (i.e. no updating), then the monopoly mechanism is individually rational relative to the status quo alternative of Cournot competition. But if the inferences are required to satisfy our credibility conditions, then the monopoly mechanism will not be unanimously ratified by all types of all players.

2.2 Minimum Quantity Restrictions

The effect of our credibility requirement is not always to strengthen the participation constraints. We demonstrate this below in the context of the same duopoly example, but with a different proposed mechanism. In particular, let δ be the decision rule associated with the (unique) equilibrium outcome from Cournot competition with a minimum quantity restriction Q (i.e. if a firm decides to produce it must produce at least Q).

Proposition 3. *For Q sufficiently small, Cournot competition with a minimum quantity restriction of Q is ratifiable against the status quo alterative with no quantity restriction, but for all $Q > 0$ such a mechanism is not individually rational with passive updating.*

Proof. See the appendix. ■

The intuition behind this example is straightforward. A high-cost firm that would produce an amount much less than Q in the status quo game, and receive a small profit, finds it unprofitable to produce with the minimum quantity restriction, and so gets zero. Such a firm would veto the minimum quantity restriction in favor of the status quo if the status quo is played with the prior beliefs. But it is credible for the ratifier to infer that vetoer has a high cost as a result of the veto, which makes the veto unprofitable. Hence, with learning no type wants to veto.

3 Alternative Formulations of Ratifiability

The issue of ratifiability is closely intertwined with the issue of defining beliefs "off the equilibrium path" in a game of incomplete information. Our definition of ratifiability challenges the notion that any beliefs, and in particular passive beliefs, can follow an unexpected veto.

In the last several years, a considerable literature on "equilibrium refinements" has built up around precisely this problem of specifying plausible beliefs off the equilibrium path. We now compare our choice of refinement and its implications for ratifiability to some alternative refinements. The purpose of this section is to motivate and clarify why we settled on a definition of ratifiability based on the refinement of Grossman and Perry (1986). The main problem with the other refinements is that they generally violate a natural "rational expectations" requirement that ratifiability satisfies: those types that are believed to have vetoed are precisely the types that

benefit from vetoing if everyone believes that they are the types to veto.

To save space, formal definitions are omitted and instead we show how these alternative refinements would apply to the duopoly problem analyzed in the previous section, with costs uniformly distributed on [0,1] and G given by Cournot competition. As in the first part of the last section, let δ be the monopoly mechanism. In this section we are interested in determining whether the monopoly mechanism is ratifiable relative to the Cournot status quo, but where ratifiability is not defined relative to "credible veto sets" but relative to veto sets that satisfy restrictions derived from alternative refinements. In particular, we investigate three refinements: the intuitive criterion, divinity, and universal divinity.

Before proceeding, it is useful to restate one property of ratifiable mechanisms in the context of our duopoly example. Recall that in the proof of Proposition 2, we showed that δ is ratifiable relative to G, so long as the veto set V is such that the highest-cost vetoer to produce is sufficiently small ($c_v(V) \leq .8515$), or equivalently, if the ratifier's expected output is sufficiently high in G. In fact, this is true regardless of which refinement concept we use to place restrictions on veto sets. Hence, δ is ratifiable against G relative to a particular refinement if the refinement allows a veto set V such that $c_v(V) \leq .8515$.

3.1 The Intuitive Criterion

The *intuitive criterion* (Cho and Kreps, 1987) requires that beliefs be concentrated on those types for whom there exists some belief such that if the ratifying firm inferred those beliefs from a veto, then this type of firm would wish to veto. In other words, no weight can be put on types for whom the deviation of vetoing is "bad" in the sense that the vetoer surely loses from the deviation (prefers the mechanism to the status quo) regardless of what belief the veto induced on the other firm.

This is a much weaker requirement than the one we proposed in the previous section. If a mechanism is ratifiable relative to the status quo, then there will always be a sequential equilibrium of the ratification game where everyone ratifies the mechanism and beliefs satisfy the intuitive criterion. In fact, for this example, there are beliefs satisfying the intuitive criterion that support the monopoly mechanism. This is demonstrated below.

We begin by calculating the *believable set*, the set of types for which there exists some belief that makes vetoing profitable. Regardless of type, the belief that makes vetoing the most profitable is $c = 0$, because it leads to the smallest production by the other firm in the status quo. Therefore, type c is believable if c does at least as well by vetoing the mechanism if the ratifying firm infers that the vetoer's cost is 0. With beliefs $V = \{0\}$, $c_v = .93$. Beliefs must be concentrated on types for whom $U(c,\{0\}) \leq 0$, which implies that the believable set is $[0, .562]$.

The belief for the ratifier satisfying the intuitive criterion that is most apt to support the monopoly mechanism is the most *optimistic* belief; namely, for the ratifier to infer that the vetoers cost is the largest believable type $c = .562$. It is easy to verify that such a belief supports the monopoly mechanism: all types prefer the monopoly outcome to the status quo when by vetoing the mechanism the vetoer reveals that its cost is .562. For beliefs $V = \{.562\}$, $c_v < .8515$, so the monopoly mechanism is ratifiable under the intuitive criterion.

The intuitive criterion is a weak refinement in this application, and similar applications with a continuum of types, because of the extreme beliefs that it allows. The believable set is defined by the ratifying firm making the *most pessimistic* inference ($c = 0$), but then the individual rationality constraint is determined by making the *most optimistic* inference ($c = .562$). The intuitive criterion allows beliefs that are far from consistent in a rational expectations sense.

3.2 Divinity

The *divinity* refinement (Banks and Sobel, 1987) goes one step further than the intuitive

criterion by imposing a monotonicity condition on beliefs in the believable set. For any two believable types, t and t', with the property that the set of beliefs under which t prefers to veto strictly contains the set of beliefs under which t' prefers to veto, divine beliefs must have a higher likelihood ratio of t to t' than the prior likelihood ratio between the two types. As in the intuitive criterion, we must find the most optimistic belief possible, but subject to the likelihood ratio constraint. It is easy to show that the most optimistic divine belief is simply the truncated prior on the believable set. In our example, this reduces to a uniform posterior on the believable set. With this belief, a veto of the mechanism will result in Cournot competition in which the vetoer believes that the ratifying firm's cost is uniformly distributed on $[0,1]$ and the ratifier believes that the vetoing firm's cost is uniform on $[0, .562]$. With $V = [0, .562]$, $c_v = .878 > .8515$, so the monopoly mechanism is not ratifiable with divine beliefs. In our example, divinity is strong enough to eliminate the monopoly outcome.

3.3 Universal Divinity

The stronger refinement of *universal divinity* (Banks and Sobel, 1987) is a strengthening of divinity that requires the likelihood ratio condition to hold relative to *any* prior, not just the original prior. This condition essentially requires that a belief places zero probability not only on types outside the believable set, but also on most other types as well. Which types must receive zero probability is determined in the following way. For every belief p concentrated on the believable set, let $C(p)$ denote the set of types who are at least as well off vetoing the monopoly mechanism and playing the Cournot game in which the ratifier has these beliefs about the vetoer's type and the vetoer has the original prior about the ratifier, and let $\bar{C}(p)$ denote the set of types who strictly prefer vetoing the monopoly mechanism when the ratifier has beliefs p in the status quo. A type c must receive zero probability in a universally divine belief if for every belief p in the believable set, if $c \in C(p)$ then there exists another type $c' \in \bar{C}(p)$.

If the payoff structure of a game possesses a monotonic structure in types, universally divine beliefs are concentrated at one point, corresponding to the type which, at least for some beliefs p in the believable set, is the unique member of $C(p)$ and, for all p' concentrated in the believable set, is in $C(p')$ whenever any other type is in $C(p')$. More generally, this criterion reduces to a set of types, which can be called the *universally divine set*. This set will include (at least) every type for which there exists some belief p concentrated in the believable set with the property that that type is the unique element of $C(p)$ and will exclude every element which is never a unique element of $C(p)$ for any p concentrated in the believable set.

In our example, universally divine beliefs are not difficult to calculate, since there is a unique type in the universally divine set; namely, $\hat{c} = 1/2 - (c_v - 3/4)^{1/2}$ where $c_v = .8515$, so $\hat{c} = .1814$. [A belief that yields $c_v = .8515$ is $V = \{.3912\}$.] It is easy to show that with beliefs concentrated on .1814, a nonempty range of types strictly prefer to veto. Therefore, the monopoly mechanism is not universally divine.

Although divinity and universal divinity are able to eliminate the monopoly mechanism in our example, we find them less appealing than our refinement, because, like the intuitive criterion, divine and universally divine beliefs are not consistent in a rational expectations sense. The set of types who benefit from vetoing is different from the set of types who are assumed to have vetoed the mechanism.

4 Conclusion

A basic tenet of most studies in mechanism design is that the parties voluntarily decide to participate in a proposed mechanism. Indeed, if participation constraints are ignored, then it typically is possible to overcome incentive problems caused by informational differences. Our

purpose in this paper has been to take a closer look at participation constraints and a party's decision to participate. Central to this goal is specifying what happens if the parties fail to agree to participate in a proposed mechanism. We consider situations where the appropriate "status quo" is a noncooperative game, rather than some constant payoff.[5] In this case, the status quo payoffs may depend on inferences based on the parties' participation decisions. Participation constraints, then, depend on the beliefs the parties hold following a veto of the proposed mechanism. This paper has proposed and illustrated a concept of ratifiability that takes into account in a consistent way the dependence of beliefs on inferences from a veto.

Appendix

Proof of Proposition 2. We first calculate the unique credible veto set V. Since the two firms are symmetric ex ante, we can drop the subscripts i and j in what follows. Furthermore, it is easy to show that for any veto set V there is a unique equilibrium σ in G, so we suppress U_i^G's dependence on $\sigma(V)$. We need to calculate the interim payoffs from the monopoly mechanism $U_i^8(c)$ and from the status quo $U_i^G(c, V)$ for every c and every V. The first is easy to compute from a formula in Cramton and Palfrey (1990):

$$U_i^8(c) = [1/8 + (1-c)^3]/6.$$

The second function is harder to derive because the strategies for the two firms differ, since each firm has different beliefs about the other.

Denote the vetoer's strategy by $q_v(c,V)$ and the ratifier's strategy by $q_r(c,V)$. In either case, the strategy depends on a single cost level at which the firm is indifferent between producing and not.

[5]Our analysis says little about situations where the status quo is not specified. In such a setting, the mechanism designer may wish to identify an institution (a status quo) that can withstand competition from alternative institutions. Our concept of ratifiability can be used to define what is meant by a status quo that can withstand competition. Namely, we can define a "stable status quo" to be a game for which no alternative mechanism can be ratified in favor of the status quo.

Let $c_v(V)$ and $c_r(V)$ be the indifference cost levels for the vetoer and ratifier respectively. Then

$$q_v(c,V) = \begin{cases} (c_v(V) - c)/2 & \text{if } c < c_v(V) \\ 0 & \text{if } c \geq c_v(V) \end{cases} \qquad q_r(c,V) = \begin{cases} (c_r(V) - c)/2 & \text{if } c < c_r(V) \\ 0 & \text{if } c \geq c_r(V) \end{cases}$$

Let $Q_v(V)$ and $Q_r(V)$ be the expected output of the vetoer and the ratifier, respectively. Then the cost levels at which a firm is indifferent between producing and not are

(Cv) $$c_v(V) = 1 - Q_r(V), \text{ and}$$

(Cr) $$c_r(V) = 1 - Q_v(V).$$

The vetoer calculates the ratifier's expected output with the prior belief, since the veto is a unilateral deviation. Thus,

(Qr) $$Q_r(V) = \int_0^1 q_r(c,V)dc = \tfrac{1}{4}c_r(V)^2.$$

Suppose V is an interval $[\ell, h]$. Then the ratifier calculates the vetoer's expected output to be

(Qv) $$Q_v(V) = \int_\ell^h \frac{q_v(c,V)}{h - \ell}dc = \begin{cases} 0 & \text{if } c_v(V) \leq \ell \\ \dfrac{(c_v(V) - \ell)^2}{4(h - \ell)} & \text{if } \ell < c_v(V) < h \\ \tfrac{1}{2}(c_v(V) - (\ell + h)/2) & \text{if } c_v(V) \geq h. \end{cases}$$

The equilibrium in the status quo with beliefs $V = [\ell, h]$ is found by solving equations (Cr), (Cv), (Qr), and (Qv) for $c_r(V)$, $c_v(V)$, $Q_r(V)$, and $Q_v(V)$. The vetoer's payoff in the status quo then is

$$U_i^G(c,V) = q_v(c,V)^2.$$

Let $U(c,V) = U_i^\delta(c) - U_i^G(c,V)$. For a veto set to be credible, it must be that

(CV) $$U(c,V) \leq 0 \text{ for all } c \in V \text{ and } U(c,V) \geq 0 \text{ for all } c \notin V.$$

First, notice that $U(c,V) > 0$ for all $c \geq c_v(V)$, so V cannot contain $c \geq c_v(V)$. For $c < c_v(V)$, $U(c,V)$ is given by the cubic equation

$$U(c,V) = (1/8 + (1 - c)^3)/6 - (c_v(V) - c)^2/4.$$

Thus, for any V, $U(c,V)$ is parameterized by the single cut-off level $c_v \in [.75, .93]$ ($c_v = .75$ for $V = \{1\}$ and $c_v = .93$ for $V = \{0\}$). It is easy to show that $U(c,V)$ is (1) continuous, (2) decreasing for c near 0, (3) has a minimum at $c = \hat{c} = 1/2 - (c_v - 3/4)^{1/2}$, (4) has a point of inflection at $c = 1/2$, and (5) has a maximum at $c = 1/2 + (c_v - 3/4)^{1/2}$. Hence, δ is individually rational relative to G if and only if V is such that c_v is sufficiently small. Calculation shows that this critical cutoff level is $c_v = .8515$. Moreover, the only candidate for a credible veto set is an interval $[\ell, h]$ around $\hat{c}$. One such veto set takes the form $V = [0, h]$, where $h < c_v(V)$ is such that $U_i^\delta(h) = U_i^G(h,V)$; that is, h solves $(1/8 + (1-h)^3)/6 = ((c_v(V) - h)/2)^2$ and $c_v(V)$ solves (Cr), (Cv), (Qr), and (Qv). These equations are satisfied when $h = .444$. In this case, $c_r(V) = .67$, $c_v(V) = .89$, $Q_r(V) = .111$, $Q_v(V) = .333$, $U_i^\delta(h) = U_i^G(h,V) = .0496$, and (CV) is satisfied, since $U(0,V) < 0$. This veto set is unique, since raising h decreases c_v, so $U(h,V) > 0$ and h will no longer want to veto. ∎

Proof of Proposition 3. We must show that for small Q there exits a credible set V which satisfies (ii) of the definition of ratifiability. Under δ, the two firms simultaneously select output levels q_1 and q_2 subject to the constraint that if $q_i > 0$ then $q_i \geq Q$. The status quo G is the same as before: Cournot competition without quantity restrictions. A unique equilibrium is characterized for each Q by two cutoff levels, ℓ and h, with $\ell \leq h$. Each firm produces

$$q(c) = \begin{cases} 0 & \text{if } c > h \\ Q & \text{if } \ell \leq c \leq h \\ Q + \tfrac{1}{2}(\ell - c) & \text{if } c < \ell, \end{cases}$$

where the cutoff levels are determined from the equations

$$Q = (h - \ell)/2, \quad \text{and} \quad \ell^2/4 = 1 - (1+Q)h.$$

If $Q = 0$, then this mechanism is simply Cournot competition with the prior beliefs, and $\ell = h =$

.83. As Q increases, both ℓ and h decrease. This means that, relative to the status quo with the original priors, no such mechanism is individually rational with passive beliefs: a firm with costs slightly below .83 makes 0 in the mechanism but would make some small positive profit under Cournot competition.

This ignores the inferences that the other firm would make if such a firm were to veto. Intuition suggests that a vetoing firm will be suspected of having high costs, thereby destroying any benefits a high cost firm could get from vetoing the minimum-quantity mechanism. In fact, this intuition is correct for sufficiently small values of Q. For small Q, one credible set is simply the highest cost type, $c = 1$. Such an inference from a veto will lead a ratifying firm to produce much more in the status quo than it would under its original priors, which makes all types (weakly) prefer the mechanism to the status quo. A $c=1$ type earns zero profits in either case, so vetoing or ratifying by such a type can be rationalized. Thus, a minimum quantity restriction is a ratifiable mechanism, even though it is not individually rational with passive updating. ■

References

Ausubel, Larry M., and Ray J. Deneckere (1987), "A Direct Mechanism Characterization of Sequential Bargaining with One-Sided Incomplete Information," Working Paper, Northwestern University.

Banks, Jeffrey S., and Joel Sobel (1987), "Equilibrium Selection in Signaling Games," *Econometrica*, 55, 647-662.

Caillaud, Bernard, and Benjamin Hermalin (1989), "The Role of Outside Considerations in the Design of Compensation Schemes," Working Paper, University of California, Berkeley.

Cho, In-Koo (1987), "A Refinement of Sequential Equilibrium," *Econometrica*, 55, 1367-1390.

Cho, In-Koo, and David M. Kreps (1987), "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179-222.

Cramton, Peter C. (1985), "Sequential Bargaining Mechanisms." In *Game Theoretic Models of Bargaining*, edited by Alvin Roth, Cambridge University Press, Cambridge, England.

Cramton, Peter C., and Thomas R. Palfrey (1990), "Cartel Enforcement with Uncertainty about Costs," *International Economic Review*, 31, 17-47.

Crawford, Vincent P. (1985), "Efficient and Durable Decision Rules: Reformulation," *Econometrica*, 53, 817-836.

Crémer, Jacques, and Richard McLean. (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247-1258.

Farrell, Joseph (1985), "Communication and Nash Equilibrium," Working Paper, MIT.

Green, Jerry R., and Jean-Jacques Laffont (1984), "Posterior Implementability in a Two-Person Decision Problem," Working Paper, Harvard University.

Green, Jerry R., and Jean-Jacques Laffont (1985), "Implementation Through a Sequential Unanimity Game," Working Paper, Harvard University.

Grossman, Sanford J., and Motty Perry (1986), "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39, 97-119.

Holmström, Bengt, and Roger B. Myerson (1983), "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51, 1799-1819.

Maskin, Eric, and Jean Tirole (1988), "The Principal-Agent Relationship with an Informed Principal, II: Common Values," Working Paper, MIT.

Maskin, Eric, and Jean Tirole (1990), "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica*, 58, 379-410.

Matthews, Steven, Masahiro Okuno-Fujiwara, and Andrew Postlewaite (1989), "Notes on Communication in Bayesian Games: Issues and Problems," Working Paper, University of Pennsylvania.

Myerson, Roger B. (1983), "Mechanism Design by an Informed Principal," *Econometrica*, 51, 1767-1797.

Myerson, Roger B., and Mark A. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 28, 265-281.

Roberts, Kevin (1985), "Cartel Behaviour and Adverse Selection," *Journal of Industrial Economics*, 33-45.

Spier, Kathryn E. (1988), "Efficient Mechanisms for Pretrial Bargaining," Working Paper, MIT.

Spulber, Daniel F. (1988), "Contingent Damages and Settlement Bargaining," Working Paper, USC.